This book collects new and original works focusing on specialised knowledge and on the ways in which it is represented, redefined, and disseminated. The methodological approaches used combine corpus linguistics, genre studies, discourse and multimodal analysis. The volume is divided into three parts. Part 1 deals with knowledge popularization and dissemination practices in medical, environmental, business, legal, and institutional discourse domains. Part 2 analyses popularisation media, including blogs, google talks, TV interviews, TV series, websites, and the press, and illustrates the discursive features of each medium, with specific attention to how specialised topics are treated. Finally, Part 3 focuses on methods and tools to transfer specialised knowledge to novices, such as language students, university students specialising in a given subject, or translators.
This book is engaging for its variety both in terms of methodological approaches and of specialised communicative settings and it will be useful for academics and PhD students working in various fields, such as Communication studies, Translation studies, Teaching, and Language studies.

Chapters by:
Adriano Ferraresi
Giuliana Elena Garzone, Paola Catenaccio
Stefania M. Maci
Anna Franca Plastina, Rosita Maglie
Erik Castello, Sara Gesuato
Antonella Napolitano, Maria Cristina Aiezza
Camille Biros, Caroline Peynaud
Sara Castagnoli, Elena Magistro
Costanza Cucchi
Michele Sala
Jekaterina Nikitina
Sonia Piotti, Amanda C. Murphy
Alida Maria Silletti
Mathilde Gaillard
Elisa Mattiello
Gianmarco Vignozzi
Francesca Coccetta
Sara Corrizzato, Valeria Franceschi
Kim Grego, Alessandra Vicentini
Angela Zottola
Maria Teresa Musacchio, Virginia Zorzi
Veronica Bonsignori
Jane Helen Johnson
Carmela Mary White
Isabel Durán-Muñoz

Lingue e Linguaggi vol. 29 - Numero Speciale 2019

Representing and Redefining Specialised Knowledge: Corpora and LSP

a cura di
Francesca Bianchi
Elena Manca
Denise Milizia

Lingue e Linguaggi

vol. 29 - Numero Speciale
2019

Università del Salento

Lingue e Linguaggi

29/2019
Numero speciale

Representing and Redefining Specialised Knowledge: Corpora and LSP

a cura di
Francesca Bianchi
Elena Manca
Denise Milizia

UNIVERSITÀ DEL SALENTO
2019

LINGUE E LINGUAGGI

Pubblicazione del Dipartimento di Studi Umanistici dell'Università del Salento. Tutti i contributi pubblicati in Lingue e Linguaggi sono stati sottoposti a double-blind peer-review.

Numero 29/2019

DIREZIONE E REDAZIONE
Dipartimento di Studi Umanistici
73100 LECCE, via Taranto, 35
tel. +39-(0)832-294401, fax +39-(0)832-249427
Copertina di Luciano Ponzio: *Ecriture* (particolare), 2007.

http://siba.unisalento.it
ISSN 2239-0367
eISSN 2239-0359 (electronic version)
http://siba-ese.unisalento.it

Indice

PART II – POPULARISATION MEDIA

PART III – TRANSFERRING SPECIALIsED KNOWLEDGE TO NOVICES

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 7-13
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p7
http://siba-ese.unisalento.it, © 2019 Università del Salento

INTRODUCTION

FRANCESCA BIANCHI,[1] ELENA MANCA,[1] DENISE MILIZIA[2]
[1]UNIVERSITÀ DEL SALENTO, [2]UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

This book collects new and original works dealing with specialised knowledge and on the ways in which it is represented, redefined, and disseminated. The chapters in this volume have been grouped into three main parts focusing on the relationship between specialised knowledge and its dissemination from different perspectives.

The first part 'Popularising and disseminating specialised knowledge' includes articles whose ultimate purpose is providing insights into the features of the popularisation practices in medical, environmental, business, legal, and institutional discourse domains. The methodological approaches used combine corpus linguistics, genre studies, discourse and multimodal analysis.

Ferraresi's chapter aims to assess the extent to which terminological density reflects the distinction between texts addressed to experts and texts addressed to non-specialist readers. It also aims to relate quantitative findings to qualitative observations on the discursive strategies adopted in the popularised genres under consideration to target different audiences. The texts selected for analysis are scientific opinions addressed to experts published by the European Food Safety Authority and their corresponding popularised versions.

Garzone and Catenaccio explore the discursive construction of medical knowledge and of the medical profession in the 19th century, combining discourse analysis and corpus linguistics. In particular, the aim of this study is twofold: first, the analysis focuses on the representation of the medical profession; then, it shifts to the episteme of the medical science as represented in three self-help medical handbooks published in the US in the latter quarter of the 19th century, Pierce's *The People's Common Sense Medical Adviser* (1883), Byrn's *The Mystery of Medicine Explained* (1887), and *Gunn and Jordan's Newest Revised Physician* (1887).

Maci's investigation focuses on the genre of scientific posters and, particularly, on those linguistic forms regarded as evidential markers. The study is based on the analysis of the verbal components of a corpus of 28 medical posters published online between 2002 and 2011. Using a text-driven approach and already existing categorisations of evidentials, this chapter aims to show the extent to which, if any, knowledge is linguistically conveyed through evidentiality in the genre of medical posters.

Plastina and Maglie explore how vague language is used in multiple forms of specialised knowledge thus contributing to triggering the ongoing debate on MMR vaccine-induced autism. A corpus-assisted approach to discourse analysis is used to identify the functions that vague language categories play in the process of knowledge production, reception and reconstruction. The study also reveals how vague language is strategically used as a technique to affect lay people's behavior and ideas on the benefits of vaccination.

Castello and Gesuato's chapter analyses aspects of the lexico-grammar of religiously oriented environmental discourse produced by Pope Francis. The authors compare the most frequent keywords and keyword clusters of the encyclical letter *Laudato Si'* against popularised updates on scientific and technological advances available on the NASA website. The methodological approach combines quantitative corpus analysis with qualitative interpretation of keywords, clusters of keywords and concordance plots. Results suggest that a lexico-phraseological description of environmental discourse can fruitfully contribute to uncovering the ideas and rhetorical goals encoded in it.

Napolitano and Aiezza analyse a corpus of environmental information about global warming and energy policies published on official governmental websites, by comparing the data and reports produced during Obama's office with the most recent communications reflecting the priorities of Trump's administration. Adopting the methodological approach of Corpus-Assisted Critical Discourse Analysis, the authors aim to show how scientific knowledge is reproduced online to support contrasting ideologies and to justify political choices.

Biros and Peynaud's chapter aims at understanding how knowledge produced by the Intergovernmental Panel on Climate Change (IPCC) is disseminated in three types of documents with various degrees of specialization – Earth Negotiation Bulletins, United Nations reports and the general-interest press – over two time periods (2009-2010 and 2014-2017). The methodological approach selected for this analysis combines the tools of corpus linguistics and discourse analysis, which puts quantitative results in context. The purpose of the analysis is to contribute to a better understanding of the role of the IPCC as a source of scientific knowledge and to track discursive variation across discourse communities and time periods.

Castagnoli and Magistro compare CSR reports translated into English from Italian with reports originally produced in English, in order to detect differences in the way corporate identity is construed and conveyed. In order to do this, they use a bilingual corpus including translated English reports and their Italian source texts, as well as comparable originals in English and Italian. The two scholars show that (im)personalisation patterns are considerably different in original and translated English CSR reports, as the tendency to reproduce Italian conventions in the English versions clearly suggests.

Cucchi focuses her research on the English versions of German and Spanish companies' corporate websites. In particular, the purpose of the study is, first, to unveil whether discursive differences exist in the English versions set up by European local companies, second, to check whether the differences may be explained with reference to Hofstede's model. The methodological approach used combines the framework of analysis of Corpus-Assisted Discourse Studies with Hofstede's theories on value orientations. Results point to a number of cultural differences which are still evident despite the ever more increasing globalising attempts.

Sala's chapter compares abstracts in print issues and those available in electronic format of the *Washington Law Review* in order to check the relationship existing between digital and print resources in specialised communication. The focus is particularly placed on knowledge dissemination (KD) resources such as the labels used to refer to the associated research article, the verbs employed to describe the type of scientific activity performed by the research article and, finally, the metadiscursive markers (such as frame and endophoric markers) used to provide interpretive guidance to readers. These resources are investigated by using qualitative and quantitative approaches.

Nikitina's chapter focuses on the linguistic practices and strategies used when dealing with issues of bioethics in the case-law of the European Court of Human Rights ("ECtHR"). Using a corpus-based approach, the analysis investigates the interaction between legal and bioethical term-related phraseological units, their structure and distributional patterns. The ultimate purpose of the study is that of researching the balance and intersection between bioethically charged phraseology and legal phraseology.

Piotti and Murphy concentrate on analysing how specialised knowledge in infographics is communicated by means of socio-semiotic, linguistic and discursive resources. In order to do this, the authors select a corpus of infographics from the World Health Organisation. Using a qualitative approach, they analyse the different modes and their interrelations in their capacity to create all three of Halliday's meaning functions; using a quantitative approach they investigate the interrelations between the cognitive dimensions arising from the corpus data and the linguistic and discursive strategies that realise them.

Silletti's chapter deals with the political slogans used during the French presidential elections campaign in 2017 and with their diffusion on the web. Using an approach inspired to theories on enunciation and French political discourse analysis, the author identifies a series of key words contained in the political slogans presented by the eleven candidates to French presidential election and analyses how these words are reused and redefined by journalists and political opponents to talk about the candidate who carries them.

The second part of the present book, "Popularisation media", focuses on blogs, google talks, TV interviews, TV series, websites, and the press. By applying corpus-assisted methods of analysis, the authors in this part of the volume illustrate the discursive features of each medium, with specific attention to how specialised topics are treated.

Gaillard investigates the blog posts of think tanks and argues that they are characterised by distinctive rhetorical strategies which fulfil the goal to influence the audience's decision-making process. These blogs allow think tank experts to express their standing, to advocate for new ideas and to react to potential criticism. They also offer a window into the organisation's work, thus contributing to its institutional image.

Mattiello investigates videos from the *Talks at Google* series, video presentations on specialised topics given by invited speakers at Google offices throughout the world. Her analysis – focussing on both verbal strategies, such as informal register, figurative language, and narratives, and non-verbal strategies, such as hand/arm gestures, head/body movements, and gaze direction – highlights that these talks diverge from more traditional academic and institutional genres, such as academic lectures and conference presentations, and show continuity with other forms of popularisation, such as TED Talks. All in all, however, they can be considered a genre on their own.

Vignozzi analyses the occurrence of idiomatic expressions and phrasal verbs in TV interviews on a range of specialised topics and observes that they are largely used by both interviewers and interviewees, regardless of topic. He suggests that these features typical of spoken register are used to smooth interaction and engage the audience. He also observes that gestures repetitively occur together with idiomatic expressions and phrasal verbs, often to indicate their figurative meanings.

Still in the realm of TV broadcasts is the chapter by **Coccetta.** By means of a recently developed multimodal concordancing tool and of a specially tagged corpus containing all the episodes of the *House M.D.* TV series, the author analyses interaction among characters and argues that Schegloff and Sacks's adiacency pair framework should be revised to accommodate not only the highly specialised dialogue of medical settings, but also different interpersonal relationships between the characters, and the complex intersemiotic interactions characterising TV film series.

Corrizzato and Franceschi investigate how theme parks are promoted on the web, by analysing the English websites of eight European theme parks. Adopting Dann's framework of analysis, they classify the top key words characterising these websites and conclude that the theme parks considered are presented as "a post-modern, hyperreal non-place, where advertising of the fictional as authentic is combined with the subsuming of the dimensions of time and space and the suspension of the laws of physics to plunge guests into fictional realms in which fun, fairy tales and thrilling adventures are essential elements".

This part ends with three chapters dealing with specialised language in the press.

By applying corpus-assisted critical discourse analysis, **Grego and Vicentini** investigate how English and Italian online newspapers present the controversial topic of euthanasia. The authors analyse the actors, ideologies, and cultural features involved in such discourse. Their results show that, despite large differences at the legislative level between the two countries, the actors and ideologies involved in assisted-death discourse are largely similar; differences however appear in the way the legal cases of British or Italian citizens are treated.

Zottola, too, applyies corpus-assisted critical discourse analysis to discuss how transgender people are presented in British popular and quality newspapers. In this chaper she shows that terminology related to transgender identities is used inconsistently and inappropriately in both types of press, which supports the hypothesis that transgender-related terminology belongs to the realm of specialised language. The author also observes differences between quality and popular press, with a tendency of the latter "to prefer terms related to the physical aspect, or to the process of transition, using both inclusive and derogatory linguistic constructions".

Finally, **Musacchio and Zorzi** investigate the translation process of scientific 'news' from English into Italian. Using a parallel/comparable corpus of newspaper, magazine and news-agency articles in English and Italian on the vaccine debate, these authors observe that "translations into Italian are mainly rewritten and transedited even when there is a major change in readership", although some relevant instances of transcreation can also be found in a specific Italian magazine, *Internazionale*. In terms of effectiveness, it is this latter magazine that best manages to convey the intended meanings to the target audience.

The third and last part in this book focuses on methods and tools to transfer specialised knowledge to novices, such as language students, university students specialising in a given subject, or translators.

Bonsignori's chapter illustrates the ways in which film clips can be used as resource material with university students studying English for Business and Economics. Her approach includes the multimodal analysis of the selected scenes, considered as a fundamental step to interpret the message in specialised contexts. The chapter thus presents the multimodal analysis of two clips – the analysis being carried out using ELAN, an annotation software that allows researchers to create and apply personalised multi-level analytical frameworks – and suggests possible applications in class.

Using a corpus of social work discourse and a small corpus of students' assessed essays on the same topic, **Johnson** suggests ways to transform evidence from a specialised corpus into class activities aimed at raising foreign language awareness in students training in the given field. Comparison between the two corpora allowed the author to assess students' language competence prior to the classroom activities and to identify areas of weakness in student competence. Next, she devised and tested in class corpus-related activities based on the corpus' wordlist, n-grams, automatically extracted semantic fields, and concordancing.

White uses a small corpus of research articles in English to sensitise post-graduate students to academic writing. The texts that comprise the corpus are collected by the individual students. The corpus is then used by the teacher to create concordances of potentially tricky words to be illustrated in class and discussed with the students. Subsiquetly, the acquisition of the grammar, lexical or rhetorical patterns observed is supported by means of targeted production exercises.

Finally, **Durán-Muñoz** illustrates the phases required to build a corpus-based ontology in a specialised domain and argues that corpora are crucial in any terminological project, as they provide relevant information that contributes not only to the identification of terms, contexts, synonyms, equivalents, but also to the enlargement of conceptual representation.

Bionotes: Francesca Bianchi holds a Degree in Translation (University of Trieste, IT), and a PhD in Linguistics (Lancaster University, UK). Her research areas include corpus linguistics, subtitling, and technologies in language learning. She has organised international conferences and seminars in applied linguistics and audiovisual translation, in and outside Lecce, and has published several papers and edited books in the area of applied linguistics.
Elena Manca is a researcher and lecturer at the Department of Humanities, University of Salento (Lecce, Italy). She holds a PhD in English Language and Translation from the University of Salento. She teaches English Language and Translation on a BA course in Linguistic Mediation (University of Salento). Her main research interests are Corpus Linguistics and the study of meaning, Corpus Linguistics and Cross-cultural Studies applied to translation and contrastive analysis, special languages with particular attention to the language of tourism, the use of new technologies in language teaching and audiovisual translation.

Denise Milizia is Associate Professor at the University of Bari Aldo Moro. She teaches English Language and Translation on two BA courses and American and English Politics and European Affairs on two MA courses, in the International Relations and European Studies programme. Her latest publications deal with the relationship between the UK and the European Union, with a special focus on the role of metaphor in European politics. Her more recent interests lie in the spread of populism in Europe and beyond, and the relationship between Europe and Russia in a critical perspective.

Authors' addresses: francesca.bianchi@unisalento.it; elena.manca@unisalento.it; denise.milizia@uniba.it

PART I Popularising and disseminating specialised knowledge

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 17-39
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p17

HOW SPECIALIZED (OR POPULARIZED)? Terminological density as a clue to text specialization in the domain of food safety

ADRIANO FERRARESI
UNIVERSITY OF BOLOGNA

Abstract – Terminological density is often mentioned as one of the main lexical features that distinguish specialized forms of communication and their popularized counterpart, i.e. discourse conveying specialized knowledge to an audience of non-specialist readers. The underlying idea is that texts addressed to experts feature a high number of terms that are expected to be understood within the target discourse community, while in popularized texts, vocabulary is chiefly drawn from the general language, and hence density of terms is lower. Though intuitively appealing, the idea that terminological density can provide quantitative evidence as to the degree of specialization or popularization of texts has seldom been investigated empirically. Taking the domain of food safety as a case in point, the aim of this article is twofold. First, it aims to assess the extent to which terminological density reflects the distinction between more specialized and more popularized texts. To do so, it describes a method to operationalize and measure terminological density building on replicable corpus-based procedures and freely available tools for vocabulary profiling. Second, in a more descriptive perspective, it aims to relate quantitative findings to qualitative observations on the discursive strategies adopted in the popularized genres under consideration to target different audiences. The texts analyzed consist of scientific opinions addressed to experts published by the European Food Safety Authority, as well as their popularized versions (factsheets, Frequently Asked Questions and news), also produced by the Authority to disseminate knowledge to the wider public. Results provide evidence that terminological density as operationalized here not only reflects the difference between specialized and popularized texts, but can also point to more subtle differences related to popularized texts' varied audiences and discursive strategies.

Keywords: specialized corpora; popularization; terminological density; vocabulary profiling.

1. Introduction

The notion that scientific or technical texts are characterized by different degrees of specialization is a central one in several applied linguistics fields, from English for Special Purposes (ESP) to vocabulary research. Text specialization (or "technicality", see e.g. Krüger 2015) is closely linked to the level of expertise in a subject matter displayed by text producers and expected in target audiences: communication between experts in a scientific or technical domain typically features a higher degree of specialization than communication between experts and non-experts. Hence the relevance of this parameter, e.g., in assessing the suitability of texts for teaching disciplinary contents (Chung, Nation 2003), or their intelligibility by ESP students (Ha, Hyland 2017).

Focusing specifically on the strategies for conveying specialized knowledge to non-experts, generally subsumed under the umbrella term of popularization strategies (Calsamiglia, van Dijck 2004), several authors have suggested that the very distinction between specialized and non-specialized texts is not a binary one, but rather forms a continuum (e.g. Myers 2003, Delavigne 2003). In this view, popularized texts occupy intermediate positions along the continuum, and may gravitate toward one pole or the other depending on the audience that they target. This can include a wide spectrum of subjects, from a public "with little or no specialized knowledge" to "highly informed laypersons who […] may already approach semi-expert status" (Krüger 2015, p. 74), and even to experts "as soon as they step outside their very limited specialism" (Myers 2003, p. 268). In the words of Hilgartner (1990, in Myers 2003, p. 270), popularization, just like specialization, is "a matter of degree".

A large body of work has been devoted to shedding light on the diverse linguistic/discursive features that characterize popularized texts, especially by comparing them to well-established genres of scientific and technical communication, such as research articles (Gotti 2014, p. 17). Such features include greater reliance on figurative language (e.g. Constantinou 2014) and on definitional/paraphrasing structures (e.g. Gotti 2011), as well as the use of less nuanced markers of modality to express certainty and probability (e.g. Bondi 2015).

The present contribution focuses on one such feature, which is often mentioned in the literature as distinguishing highly specialized forms of communication from their popularized counterparts, i.e. terminological density (see Section 2). The underlying idea is that texts addressed to experts display a high density of domain-specific terms which are expected to be understood by the target audience, while in popularized texts vocabulary is likely to be chiefly "drawn from the general language" (Gotti 2011, p. 207), which makes density of terms lower. Though intuitively appealing, the idea that terminological density can provide quantitative evidence as to the degree of specialization or popularization of texts has seldom been investigated systematically.

Taking the domain of food safety as a case in point, the aim of this article is twofold. First, it aims to assess the extent to which terminological density reflects the distinction between more specialized and more popularized texts.

To do so, it describes a method to operationalize and measure terminological density building on replicable corpus-based procedures and freely available tools for vocabulary profiling (Nation, Heatley 2002, Anthony 2014). Second, in a more descriptive perspective, it aims to relate quantitative findings to qualitative observations on the discursive strategies adopted in the popularized genres under consideration to target different audiences. The texts analyzed consist of scientific opinions addressed to experts published by the European Food Safety Authority (EFSA),[1] as well as their popularized versions (factsheets, Frequently Asked Questions and news), also produced by EFSA to disseminate knowledge to the wider public.

The remainder of the article is structured as follows. Section 2 reviews previous work focusing on terminological density as an indicator of text specialization, and expands on the theoretical and methodological underpinnings of the notion of term adopted in the study. Section 3 presents the study methodology, and specifically the steps taken to build the corpora of texts in the domain of food safety, the method employed to extract and classify terms, and the setup of the vocabulary profiling tools used in the analysis of EFSA texts. Section 4 presents the results of the quantitative and qualitative analysis, and Section 5 concludes by discussing implications of the findings for research on broadly conceived popularization discourse.

2. Background

2.1. Text specialization and terminological density

The degree to which texts make use of domain-specific terminology is arguably among the most frequently adopted parameters to assess text specialization. Both Pearson (1998, p. 62) and Corpas Pastor and Seghiri (2010) use it to select appropriate scientific and technical texts for inclusion in specialized corpora, on the grounds that "[d]ocuments with a high level of technicality [display] 'terminological density', or in other words, a high number of units that convey specialized knowledge" (Corpas Pastor, Seghiri 2010, p. 124). Along similar lines, Cabré (2010, p. 358) suggests that density of terms can be used as a parameter to classify specialized texts, arguing that "terminological density of texts varies according to their level of specialization: [the] more specialized the text, the more terminology it will have". These studies, however, do not report on quantitative measures of term density, nor do they mention explicit criteria through which the notion could be operationalized.

Andersen (1996) and Skibitska (2015) present more quantitatively-oriented approaches to assess terminological density. The former focuses on

[1] http://www.efsa.europa.eu

operating manuals related to oil-drilling procedures distinguished by different degrees of technical complexity, while the latter compares tourist texts addressed to professionals of the tourism industry and to tourists. Both studies measure term density as the proportion of domain-specific terms to the overall number of words in texts, and find support for the hypothesis that more specialized texts, i.e. texts concerning more technical procedures in the case of Andersen (1996) or addressed to experts in the case of Skibitska (2015), are associated with a higher density of terms compared to less specialized ones. Though more systematic than the above-mentioned studies, neither Andersen nor Skibitska provide details as to the methodology adopted for term identification, raising issues as to the replicability of findings.

Perhaps the most explicit attempt at establishing a correlation between terminological density and text specialization is Piscopiello and Bertaccini's (2009) study comparing medical texts at different levels of specialization/popularization (texts addressed to experts, to medicine students and to the general public) in two languages (English and Italian). Their operationalization of terminological density corresponds to the type/token ratio measure (TTR), i.e. the proportion between the number of different words (types) and the total number of words (tokens) in a text. Results indicate that TTR only reflects a coarse-grained distinction between the most specialized and most popularized texts, and does so more reliably in Italian than in English. Going beyond the partial inconsistency of results, one wonders whether a measure like TTR, which is entirely agnostic to whether a word constitutes a term or not, can appropriately be defined as a measure of terminological density.

The method described in this study (Section 3) aims at overcoming these limitations by proposing replicable, quantitative procedures for term identification, based on present-day terminology theory and practice. This makes the object of the next Section.

2.2. Term specialization and corpus-based terminology research

The advent of descriptive approaches to terminology was a major breakthrough in terminology research (Cabré 2003). Contrary to the traditional view of prescriptive terminology, descriptive terminology holds that specialized units of knowledge (concepts) can be expressed by more than one linguistic sign (term), and that it is essential for terminology as a discipline to account for such variation. By observing the use of terms in specialized contexts, it has been shown that term variation is pervasive, and that it is motivated by cognitive, linguistic and situational factors (Freixa 2006).

The level of subject-field expertise of text producers and addressees features prominently among these factors. According to Delavigne (2003) popularized texts addressed to a lay audience may not only feature few specialized terms (see also Section 2.1), but terms themselves may display a low degree of specialization: in her example, "rougeur de la peau" (*skin redness*) and its specialized counterpart "érythème cutané" (*skin erythema*) should be treated as equivalent terms distinguished by a different level of specialization, and hence of intelligibility by the audience (Delavigne 2003, p. 89). The point is taken up by Freixa (2006, p. 59), who states that:

> [o]ne of the most interesting conclusions one can draw from the analysis of the discourse of scientific popularization is that synonymy (including reformulations, paraphrases, etc.) is used to make a non-specialized audience understand the new concepts resulting from new discoveries. Different denominations are used, intended to help the reader to understand the limits of the concept […].

In order to account for such instances of functionally-motivated terminological variation, several authors have developed procedures to identify terms characterized by different degrees of specialization, notably by relying on corpora and corpus-related methods. Most of these procedures are based on comparisons between specialized and non-specialized (reference) corpora (see Marín 2014 for an overview), on the assumption that lexical units that are salient/typical in the specialized discourse under consideration tend to have significantly higher frequency in the former than in the latter. The resulting lists of lexical units are then subjected to further analysis for term classification.

The methods developed by Kwary (2011) and Ha and Hyland (2017), both concerned with financial terminology, involve the use of corpus-external resources. The former draws on input by a professional lexicographer and a subject-field expert, who are asked to classify lexical units as domain-specific financial terms, as terms belonging in other subject-fields, or as "non-terms" belonging in the general language (see also Chung, Nation 2003). The latter proposes a five-point technicality scale, ranging from "least technical" (e.g. "fee") to "most technical" (e.g. "subprime"), and assigns terms to one category or another based on manual analysis of their meaning as codified in general language and special language lexicographic resources, as well as general-language corpora. In both cases, (non-)availability of resources in the domain under investigation, be they experts or dictionaries/glossaries, may be a major hindrance, especially when working with under-resourced languages or domains.

A fully corpus-based, domain-independent method is the one developed by Marín (2016). The author aims to categorize "sub-technical" terms, i.e. terms "common to both the general and the specialized fields or amongst scientific disciplines" (Marín 2016, p. 84), along a continuous scale of specialization, with positive/negative numerical values indicating greater/lower specialization. The classification method itself consists of an algorithm built around Williams' (1998) lexical network model, according to which "the more specialised a sub-technical term [is], the greater its capacity to attract other words in a specialised context, generating more populated networks around it and its collocates" (Marín 2016, p. 94). While firmly rooted in linguistic theory and returning promising results in terms of classification precision, Marín's method has the drawback of involving complex data processing procedures, which may be especially cumbersome for less technically-minded users.

The method for term extraction described in Section 3.3 is less sophisticated than Marin's (2016) or Ha and Hyland's (2017), and affords less precise observations on the status of terms as more/less specialized. However, it does offer the advantage of being easily implemented through freely available, user-friendly software tools. Furthermore, as will become clearer in Section 3.3, its objective is not that of maximising precision in the classification of single terms, but rather that of providing *lists* of terms that, when used in combination with vocabulary profiling tools, can provide reliable indications as to the status of whole texts as specialized/popularized.

3. Study setup

3.1. Research questions

As mentioned in Section 1, the present study aims to assess whether and how terminological density reflects the distinction between more specialized and more popularized texts. Unlike most previous studies (see Section 2.1), the notion is operationalized adopting replicable corpus-based procedures and freely available, user-friendly software tools both in the identification of domain-specific terms, and in measuring density of such terms in texts.
Taking the domain of food safety as case in point, our research questions are as follows: 1) do specialized and popularized texts display different levels of terminological density at the quantitative level? 2) Can quantitative differences (if any) be explained in terms of discursive strategies adopted in popularized texts to disseminate knowledge to a wide public?

3.2. Corpus resources

Two distinct corpus resources were built for the purposes of this study. The first one consists in the study corpus, i.e. a collection of specialized and popularized texts published by EFSA, which made the object of the quantitative and qualitative analysis presented in Section 4. The second one is a purpose-built benchmark corpus, which was used as a basis for extracting a list of terms in the domain of food safety (see Section 3.3). The two corpora are presented in what follows.

3.2.1. The EFSA corpus

The main goal of EFSA, as stated on its website, is to produce "scientific opinions and advice that form the basis for European policies and legislation"[2] in the field of food safety and related issues, such as animal health and plant protection. This research-driven, scientific role goes hand in hand with another key part of EFSA's mandate, which is to "communicate clearly […] to the public at large, to help bridge the gap between science and the consumer".[2]
The EFSA corpus includes samples of two types of texts, which reflect the two categories of stakeholders targeted by EFSA (see also Bondi 2015 on EFSA's documentation). The first are samples of scientific opinions,[3] which are "forms of expert-to-expert communication, [whose] intended addressees are mainly the scientific community and economic agents submitting applications for regulated products" (Bondi 2015, p. 95). The second are popularized versions of these texts aimed at a wider public, which can take three main forms:

- "News": these are the most frequent popularized texts on EFSA's website. They "act as key documents for users who are […] interested in checking on the final assessment provided by the Agency on the issues involved", but may also "act as 'trailers' for experts who might then want to go on and read the Opinion document" (Bondi 2015, p. 95).
- "Frequently Asked Questions" (FAQs): these texts have the typical structure of FAQs, i.e. they are lists of questions and answers on topics that are supposed to be of particular relevance to the public, such as chemicals in food or dietary reference values. Not all scientific opinions are accompanied by FAQs.
- "Factsheets": factsheets are summarized versions of scientific opinions targeted to the general public. Unlike FAQs and news, which are published

[2] http://www.efsa.europa.eu/en/aboutefsa
[3] EFSA also provides other types of scientific and technical documents, such as statements by the Agency or scientific and technical reports. These are labelled as "other scientific outputs" or "supporting publications", and are less central to EFSA's mission (see http://www.efsa.europa.eu/en/efsajournal/scdocdefinitions). For this reason, only scientific opinions were taken into account.

as text-only web pages, factsheets are published in PDF format, and include images and tables. Only few factsheets have been published, on a limited range of topics; no information is made available by EFSA on the criteria that were adopted for the selection of topics, which range from substances like caffeine or aspartame to medical issues related to food-borne diseases.

The point of departure for corpus compilation was the list of "topics" available on the EFSA website:[4] each document published by EFSA is associated with one such topic, and, in turn, each topic is associated with at least one macro-category. For each topic that was labelled as pertaining to the macro-category of food safety (as opposed to, e.g., biological hazards), a manual check was performed on the number and types of documents available. Only topics for which EFSA had published at least one scientific opinion and at least three popularized texts (belonging to one of the three categories mentioned above) were retained. All available FAQs and factsheets on these topics were downloaded, plus a maximum of two scientific opinions and ten news per topic. Since the measures of terminological density adopted (see Section 3.3) are sensitive to extreme differences in document length, news texts shorter than 200 words were discarded, while scientific opinions, which tend to be very lengthy documents, were cut to a maximum of 6,200 words, corresponding to 1.5 times the size of the longest popularized text. Table 1 provides information on the resulting corpus and its sub-corpora.

	Factsheets	FAQs	News	Scientific opinions	TOTAL
Texts	7	15	191	60	273
Tokens	11,446	33,285	106,032	319,040	469,803
Types	1,778	3,153	5,841	14,532	25,304
Text length: mean	1,638	2,221	556	5,385	–
Text length: std. deviation	400	1,240	279	412	–
EFSA topics (examples)	Acrylamide, Aspartame, Bisphenol A, Caffeine, Chemicals				

Table 1
Size information about the EFSA corpus.

3.2.2. *The food safety benchmark corpus*

The aim of the benchmark corpus was to provide an "external" resource (with respect to the study corpus) of specialized texts from which food safety terms could be extracted. Research articles from the *Journal of food safety*,[5] published by Wiley, and the *Journal of food protection*,[6] the official journal of the *International Association for food protection*, were included.

[4] http://www.efsa.europa.eu/en/topics-complete-list
[5] https://onlinelibrary.wiley.com/journal/17454565
[6] http://jfoodprotection.org/

To maximize comparability between the study corpus and the benchmark one, for each topic represented in the EFSA corpus a search was performed via Google in the databases of the two journals, using the name of the topic itself (see examples in Table 1) as a query term. A maximum of 10 articles per topic per journal were downloaded. Since texts were only used for term extraction and not for comparison with the EFSA texts, differences in text length were not considered as relevant, and hence whole texts were retained. Information on the resulting corpus can be found in Table 2.

Texts	334
Tokens	2,359,388
Types	73,980
Sources	*Journal of food safety, Journal of food protection*

Table 2
Basic statistics on the food safety benchmark corpus.

3.3. Software tools and vocabulary lists

The method proposed in this study to measure terminological density is based on so-called vocabulary profiling (VP) software. Initially developed for language teaching purposes, and especially to assess intelligibility of texts by students of English for Academic Purposes, VP tools use lists of words matching pre-defined criteria, for instance featuring high-frequency words with which students are expected to be familiar vs. low-frequency words which are less likely to be known to them (e.g. Webb, Nation 2008). Based on such lists, VP tools can then measure the density of the (categories of) words of interest in the texts to be analyzed. Density itself is expressed as a proportion, e.g. of high-frequency vocabulary to the overall number of tokens or types in texts.

The study presented in Section 4 draws on a combination of existing and purpose-built vocabulary lists, representing general and specialized language. To assess the degree to which EFSA texts rely on general, high-frequency vocabulary, the *General Service List* (GSL, West 1953) was used. The GSL was complemented by a set of specialized term lists, and specifically: *a)* the *Academic Vocabulary List* (AVL, Gardner, Davies 2014), a genre-specific list of terms that are expected to be typical of scientific/academic writing in general, irrespective of differences among subject fields; and *b)* a domain-specific term list, featuring terms that are expected to be typical of the sub-

domain of food safety, created specifically for the study with the help of the CasualConc concordancer.[7]
Following common practice in terminology extraction (see Section 2.2), the list of food safety terms was created on the basis of a statistical comparison between the food safety benchmark corpus (see Section 3.2.2) and a general corpus of English, i.e. a 17-million-word subset of ukWaC (Baroni *et al.* 2009). Terms were defined as single-word lexical units occurring with a significantly higher frequency in the food safety corpus than in the general language corpus.[8]

Further, to reflect the view that domain-specific terms can differ in their degree of specialization (see Section 2.2), a set of parameters were taken into account to categorize them as "core" vs. "non-core" terms, corresponding to less vs. more specialized lexical units. Core terms were defined as those units that were distributed across at least 30% of the benchmark corpus texts (a parameter adopted, among others, by Coxhead (2000) and Marín (2016)), or whose frequency comparison with respect to the ukWaC corpus displayed a low effect size,[9] pointing to the fact that they tended to also be frequent in general language. By contrast, non-core terms were those which occurred in less than 30% of the benchmark corpus texts, or whose frequency comparison displayed a high effect size. Table 3 presents examples of the two types of terms.

Core terms	**Non-core terms**
food, see, salmonella, coli, samples	campylobacter, inactivation, STEC, ND, jejuni

Table 3
Top 5 core and non-core food safety terms, in decreasing frequency order.

Being based on single lexical units and simple frequency/distributional information, this method has obvious disadvantages: it does not take into account multi-word terms, and disregards refined semantic distinctions between specialized and non-specialized senses of words. However, as also argued in Section 2.2, it was not conceived as a full-fledged terminology classification method, but rather as a viable, easy-to-implement alternative to more sophisticated classification systems see Section 2.2. We leave it to further

[7] https://sites.google.com/site/casualconc/. CasualConc was the concordance of choice since it provides more advanced functionalities compared to other freely available concordancers. E.g. it computes word distribution values across corpus texts as well as other useful statistics, such as effect size measures for keywords (see Footnote 9).
[8] Specifically, terms were defined as lexical units for which the comparison returned a Log-Likelihood value equal or greater than 3.84, corresponding to a significance level of $p < 0.05$ (Rayson, Garside 2000).
[9] Effect size was assessed using the %DIFF measure (Gabrielatos, Marchi 2012). Low/high %DIFF values were defined as values below/above the median of %DIFF values for all terms.

work to experiment with the output of these systems as sources of data for terminological density measurements.

In the final stages of the data preparation phase, the four word lists (i.e. the GSL, AVL, and the core and non-core food safety term lists) were loaded in the VP tool AntWord Profiler (Anthony 2014). The four EFSA sub-corpora were then also loaded, one at a time, to measure their terminological density. This was calculated on a text-by-text basis, and defined as the proportion of specialized terms to the overall number of words per text; in order to reduce the effect of text length, analyses were carried out taking into account word types rather than tokens. Figure 1 displays the AntWord Profiler settings adopted.

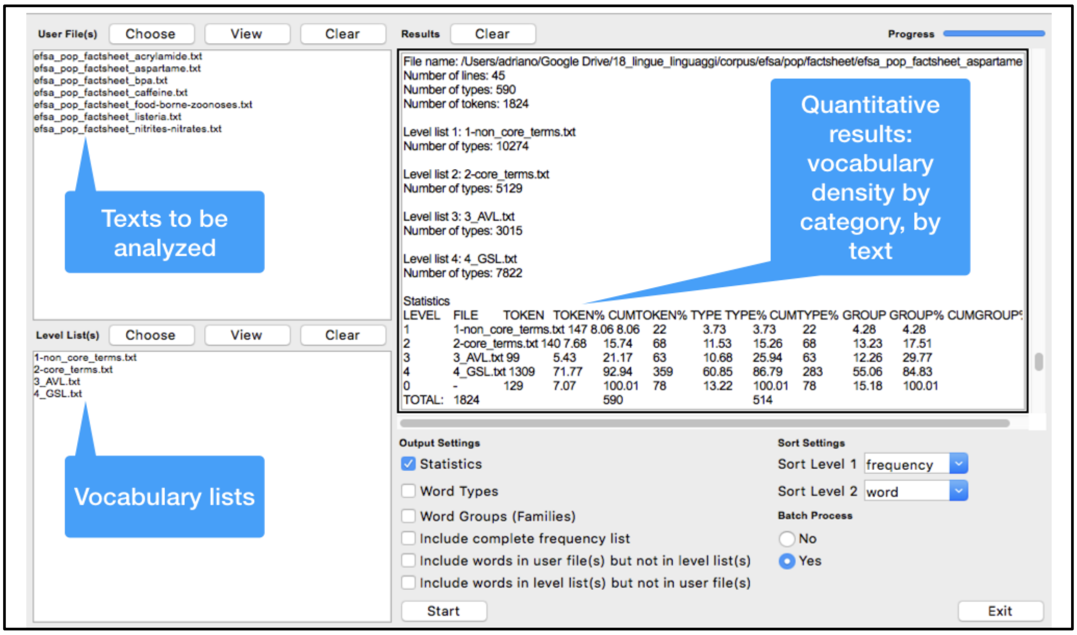


Figure 1
The AntWord Profiler interface.

3.4. Statistical procedures for the quantitative analysis

Results obtained from AntWord Profiler were exported to a spreadsheet, and statistical analyses were carried out to detect significant differences in terms of terminological density among sub-corpora. Analyses were performed in R.[10] Kruskal-Wallis tests were used to compare all sub-corpora. In the cases where they returned significant results, they were followed by post-hoc comparisons between pairs of sub-corpora using Wilcoxon rank sum tests with Holm correction. Results of these quantitative analyses are presented in Section 4.1.

[10] https://www.r-project.org

4. Results

4.1. Quantitative analysis: comparing terminological density across specialized and popularized texts

4.1.1. Overall analysis

To provide a general overview of the data, a mosaic plot is presented in Figure 2 showing the overall vocabulary profiles of the 4 EFSA sub-corpora. The height of the coloured bars indicates the density, expressed as a percentage, of three categories of vocabulary, i.e. *a)* specialized vocabulary, which corresponds to the density of genre- and domain-specific terms (see Section 3.3); *b)* general vocabulary, corresponding to the density of high-frequency words included in the GSL; and c) "other" vocabulary, i.e. a miscellaneous category of words which were not attested in any of the available lists, and which includes numbers, low-frequency words, words not typical of the domain, etc..

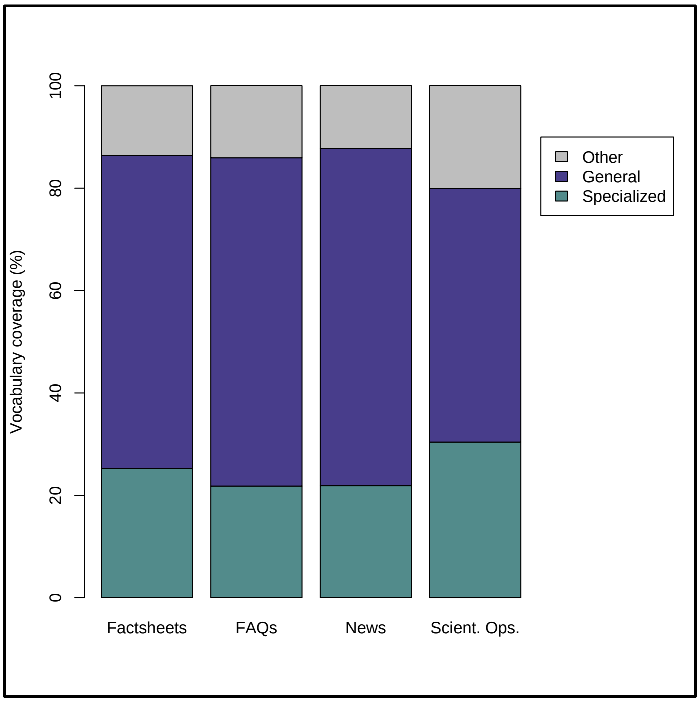


Figure 2
Overall vocabulary profiles of EFSA texts.

Focusing on the first two categories, FAQs and news are very similar in terms of density of both specialized and general vocabulary. Term density in factsheets is slightly higher, but, overall, the three popularized genres are more similar to each other than they are to scientific opinions, which display the highest density of terms and the lowest density of general vocabulary.

Zooming in on the parameter that is most central to the analysis, i.e. density of terms, the results of the statistical tests confirm the observations made on the basis of data visualization. A Kruskal-Wallis test reveals that differences among the four sub-corpora are significant (Kruskal-Wallis $\chi^2(3)=118.35$, $p < 0.001$), and results of post-hoc comparisons (see Tables 4a and 4b) show that terminological density is significantly higher in scientific opinions than in any of the popularized texts. By contrast, none of the comparisons between popularized texts display significant differences.

	Median (%)	SD
Factsheets	25.0	2.8
FAQs	22.1	3.0
News	21.9	4.1
Scient. Ops.	29.9	3.2

	Factsheets	FAQs	News
FAQs	0.051	-	-
News	0.051	0.946	-
Scient. Ops.	**0.002**	**<0.001**	**<0.001**

Table 4a; Table 4b
Terminological density; *p*-values of pairwise comparisons (significant results in bold).

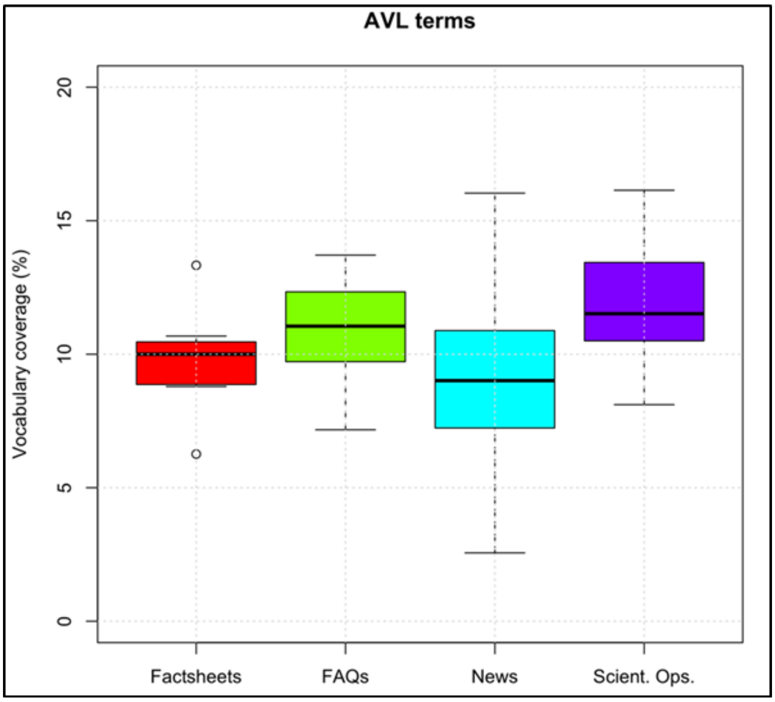


Figure 3
Density of academic terms in EFSA texts.

4.1.2. Detailed analysis

As a follow-up to the overall analysis, the density of the different categories of terms was also considered. Graphical representations are provided in this case in the form of boxplots. For term density in the texts of each sub-corpus, these show the central tendency (median values, visualized as a thick black line cutting the "boxes" in two), the spread of the data around it (corresponding to the size of the boxes; the larger the box, the more different the texts within the sub-corpora), and the minimum and maximum values of the distribution (represented by the extreme points above or below the boxes).

Starting from the academic (AVL) terms (Figure 3) the comparison among sub-corpora returns significant results (Kruskal-Wallis $\chi^2(3)=54.2$, $p < 0.001$); post-hoc tests (see Tables 5a and 5b) reveal instead a different picture. As expected, scientific opinions display the highest density of academic terms. This is significantly higher than the density in factsheets and news, but not significantly higher than the density in FAQs, which in turn are significantly denser than news (but not factsheets). In this case, popularized texts do not form a homogenous group, and this is especially due to the high density of academic terms in FAQs.

	Median (%)	SD
Factsheets	10.0	2.2
FAQs	11.1	1.9
News	9.0	2.5
Scient. Ops.	11.5	2.0

	Factsheets	FAQs	News
FAQs	0.441	-	-
News	0.517	**0.036**	-
Scient. Ops.	**0.043**	0.441	**<0.001**

Table 5a; Table 5b
Density of academic terms; *p*-values
of pairwise comparisons (significant results in bold).

The results for the core terms analysis are displayed in Figure 4. Scientific opinions are, again, the terminologically densest texts. Unlike in the case of academic terms, here factsheets are the odd ones out among popularized texts. Their term density is not significantly different from that of scientific opinions, but is significantly higher than that of both FAQs and news (Kruskal-Wallis $\chi^2(3)=73.6$, $p < 0.001$; results of post-hoc tests in Tables 6a and 6b). In other words, when it comes to density of less specialized terms related to food safety, the EFSA texts cluster in two groups, with scientific opinions and factsheets being significantly denser than news and FAQs.

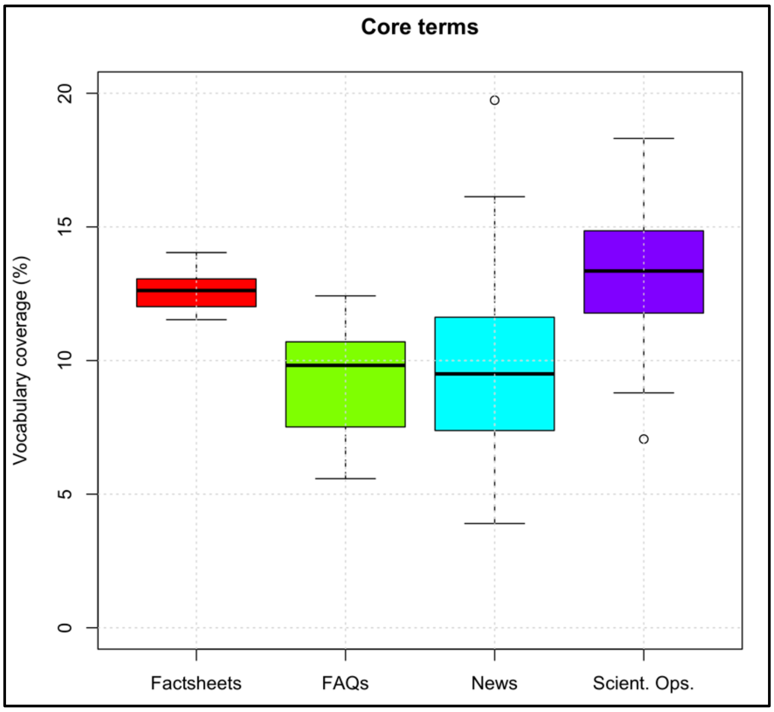


Figure 4
Density of core terms in EFSA texts.

	Median (%)	SD
Factsheets	12.6	0.9
FAQs	9.8	2.0
News	9.5	2.9
Scient. Ops.	13.4	2.1

	Factsheets	FAQs	News
FAQs	**<0.001**	-	-
News	**0.008**	0.551	-
Scient. Ops.	0.430	**<0.001**	**<0.001**

Table 6a; Table 6b
Density of core terms; *p*-values
of pairwise comparisons (significant results in bold).

We finally consider non-core terms (Figure 5), corresponding to the most specialized domain-specific terms. The pattern here resembles the overall one presented in Section 4.1.1, with scientific opinions being significantly denser than any of the popularized texts (Kruskal-Wallis $\chi^2(3)= 53.8$, $p < 0.001$; results of post-hoc tests in Tables 7a and 7b). Among these, however, the news sub-corpus is significantly denser than the sub-corpus of FAQs (but not denser than that of factsheets). This is probably due to the relatively large dispersion of data (see the large light blue box in Figure 5, as well as the SD value in Table 7a), pointing to the fact that news constitute a less homogeneous group than the rest of the popularized texts: some of the news texts have a density of non-core terms that is comparable to that of the densest scientific opinions. It should be noticed that such relative lack of homogeneity is also testified by the analyses of academic and core terms: SD values pertaining to news are consistently the highest ones among popularized texts.

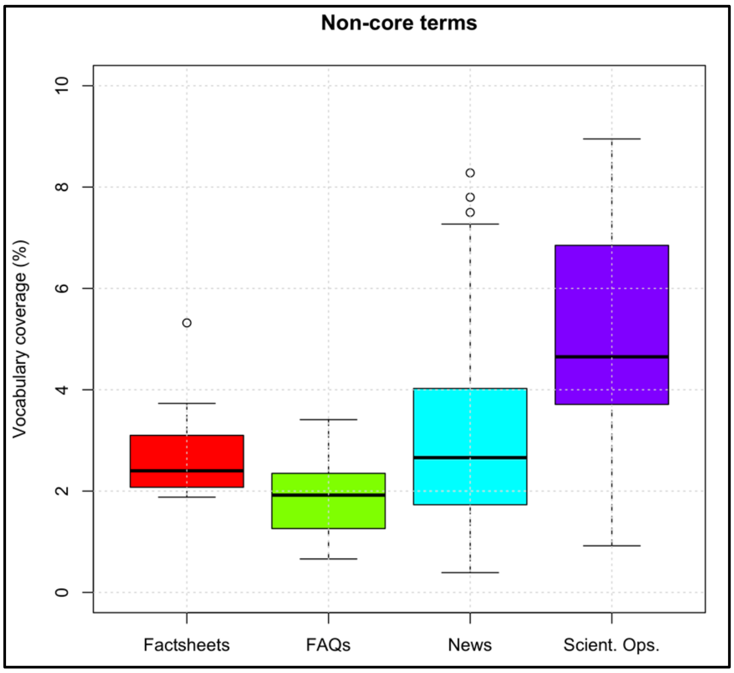


Figure 5
Density of non-core terms in EFSA texts.

	Median (%)	SD
Factsheets	2.4	1.2
FAQs	1.9	0.8
News	2.7	1.7
Scient. Ops.	4.7	2.1

	Factsheets	FAQs	News
FAQs	0.132	-	-
News	0.976	**0.028**	-
Scient. Ops.	**0.028**	**<0.001**	**<0.001**

Table 7a; Table 7b
Density of non-core terms; *p*-values
of pairwise comparisons (significant results in bold).

Summing up, differences in terminological density were observed not only between specialized and popularized texts, but also between the different genres of popularization; such differences are related to the different types of terms considered, such that different genres are characterized by relative overuse of specific categories of terms. The next Section aims at explaining the quantitative trends observed here by adopting a more qualitative perspective, taking a closer look at discursive patterns surrounding the use of terms in popularized texts.

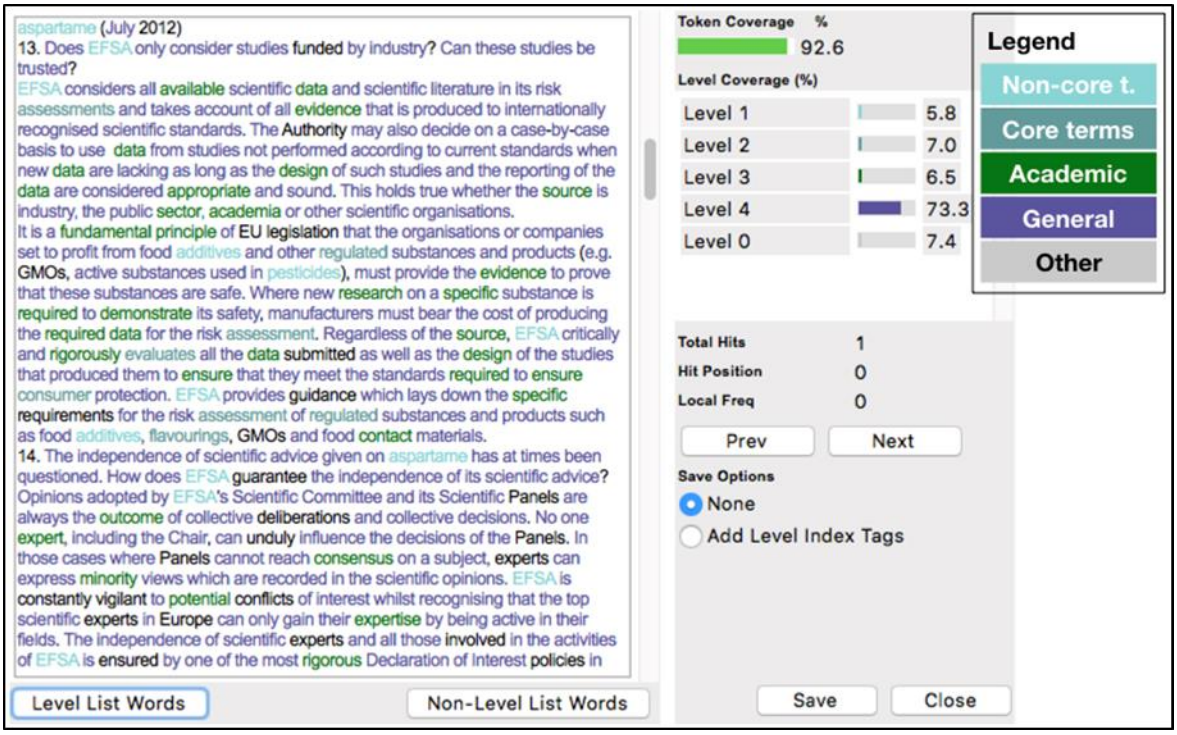


Figure 6
The text analysis function in AntWord Profiler.

4.2. Qualitative analysis: exploring terminological density in popularized texts

In order to look for possible explanations of the patterns of terminological density observed in popularized texts (see Section 4.1), we will consider here single texts belonging to different types of popularized genres and displaying especially high values of density of terms. To do so, we will rely on another feature of the AntWord Profiler tool (see Section 3.3), which makes it possible to visualize, through colours, which words were assigned to which vocabulary category in a text. Figure 6 illustrates this visualization function.

Starting from FAQs, a possible reason for their relatively high density of academic vocabulary is the emphasis they place not only on the results of EFSA's research initiatives on specific topics, but also on the methods and the wider scientific context of research itself. Examples 1 and 2 below, taken from FAQs on endocrine active substances and aspartame respectively, illustrate the point.[11] Underlined words correspond to terms classified in AVL as typical of academic registers:

[11] If not specified otherwise, examples are selected from the five texts displaying the highest density values for the category of terms under consideration.

1) On 4 July 2017 EU Member States voted in favour of the European Commission's proposal on scientific criteria to identify endocrine disruptors in the field of plant protection products. The criteria identify known and presumed endocrine disruptors. They also specify that the identification of an endocrine disruptor should be carried out by taking into account all relevant scientific evidence including animal, in vitro or in silico studies, and using a weight-of-evidence approach.
2) Where new research on a specific substance is required to demonstrate its safety, manufacturers must bear the cost of producing the required data for the risk assessment. Regardless of the source, EFSA critically and rigorously evaluates all the data submitted as well as the design of the studies that produced them […].

In these examples, the aim of EFSA is not that of explaining whether or why endocrine active substances or aspartame may be dangerous to consumers' health; rather, it aims at describing the working methods and the guiding principles that were adopted in carrying out research, possibly addressing an audience of scientists or expert legislators (see the first sentence of Example 1). It should be noticed, however, that FAQs are composite texts made of several questions and answers, each potentially targeted to a different audience. In this light, the high density of academic terms, but not, say, of domain-specific terms (see Section 4.1.2), could be due to the fact that these popularized texts are targeted to a mixed audience, which may include academics and other experts in addition to a more "traditional" audience of lay persons.

Moving on to factsheets, it will be remembered that this sub–corpus displays a significantly higher density of core (less specialized) terms than the other popularized sub-corpora. Analysis of the top 5 texts by term density reveals that core terms are often either accompanied by definitions/paraphrases explaining them (see Example 3; core terms underlined), or may be part of the explanations themselves (Example 4).

3) Listeria is a group of bacteria that contains ten species. One of these, Listeria monocytogenes, causes the disease "listeriosis" in humans and animals.
4) Exposure – concentration or amount of a particular substance that is taken in by an individual […] in a specific frequency over a certain amount of time. When experts assess consumers' dietary exposure to a chemical substance, they combine data on its concentrations in food with the quantity of those foods consumed.

In Example 3 and 4, the core terms "Listeria" and "exposure" are clearly thought to be in need of an explanation, which in the former case is provided through an in-text definition ("Listeria is a group of bacteria that contains ten species"); in the latter case, "exposure", a recurrent term in the text from which the example is taken, i.e. a factsheet on food nitrites and nitrates, is defined within a glossary (a strategy that is adopted for selected terms in all the factsheets considered). In turn, the definitions provided rely on other core terms. Factsheets thus seem to target a more homogeneous audience than

FAQs: the frequent use of terms characterized by a low degree of specialization or accompanied by explanations/paraphrases, paired with a non-significant overuse of the most specialized terms (see Section 4.1.2), seem to testify to the fact that the main audience of factsheets is mainly one of non-experts.

The same cannot be said of news, the last category of texts considered. In Section 4.1.2, the news sub-corpus was revealed as the one characterized by the highest degree of variation, especially in the analysis pertaining to non-core, highly specialized terms. Hence, part of the news texts display a degree of term density that is comparable to that of other popularized genres; other news texts, however, display peaks of density that make them more similar to specialized texts than to popularized ones. Closer inspection of the former set of texts shows that, like in the case of factsheets, highly specialized terms, if present, are accompanied by definitions or paraphrases (see Example 5; non-core terms underlined).

5) At the end of last year the European Food Safety Authority (EFSA) published its first annual report[1] on infectious diseases <u>transmissible</u> from animals to humans (<u>zoonoses</u>) which affect over 380,000 EU citizens per year.

By contrast, the news texts displaying high term density values seem to reproduce the style and content of the scientific opinions that they should contribute to disseminate (see Examples 5 and 6).

6) Bacteria that produce <u>ESBL</u> <u>enzymes</u> show multi-drug resistance to β-lactam <u>antibiotics</u>, which include <u>penicillin</u> <u>derivatives</u> and <u>cephalosporins</u>. The prevalence of <u>ESBL</u>-producing E. coli varied across countries, from low to very high.
7) The report also provides data on other important <u>zoonotic</u> diseases such as Escherichia coli (<u>verocytotoxin</u> and <u>Shiga</u> toxin producing E. coli, collectively termed <u>VTEC</u>), <u>Mycobacterium</u> bovis, <u>Brucella</u>, <u>Yersinia</u>, <u>Trichinella</u>, Echinococcus, and <u>Toxoplasma</u>.

This is not to say that no attempt is made in these examples at paraphrasing or expanding on the most technical notions (e.g. “β-lactam antibiotics […] include penicillin derivatives and cephalosporins”, and “verocytotoxin and Shiga toxin producing E. coli, collectively termed VTEC”). Nonetheless, the degree of domain expertise expected on the part the reader is arguably higher than in factsheets (e.g. readers are expected to understand what “cephalosporins” are). The accumulation of several highly-specialized terms (“Brucella, Yersinia, Trichinella”) further contributes to the impression that these news items are targeted to an expert, rather than lay, audience (see also Bondi 2015, p. 95).

The next Section summarizes and discusses these results.

5. Conclusion

This article set out to test empirically the hypothesis that terminological density can be used as a measurable indicator to tell apart specialized and popularized texts, and that quantitative differences in terms of this parameter can be related to differences in terms of discursive strategies and target audiences of the texts considered. Focusing on the documentation published by EFSA on its website, it compared specialized scientific opinions and three different popularized genres, namely factsheets, FAQs and news. To do so, it proposed a method to operationalize and measure terminological density that is entirely based on replicable procedures, and which can be implemented using freely-available, user-friendly software tools.

Results of the quantitative analyses provided evidence that terminological density as operationalized here does reflect the difference between specialized and popularized texts, and can further point to more subtle differences related to popularized texts' varied audiences and discursive strategies. Analyses showed that scientific opinions generally display a higher density of terms than their popularized counterparts. At the same time, popularized texts were revealed as a non-homogeneous group, providing support for the view that popularization, on a par with specialization, should be conceived as a multi-faceted and gradable phenomenon (see Section 1).

Specifically, FAQs and news were found to display varying degrees of specialization (or popularization, depending on the perspective adopted), by combining popularizing strategies and a relative lack thereof: the former make very limited use of domain-specific terms and instead rely extensively on academic vocabulary (at least in certain text parts); the latter include texts where highly specialized food safety terms are accompanied by definitions and paraphrases, and other texts which closely reproduce the contents and style of specialized scientific opinions. Such differences, it was argued, might be related to the mixed audiences of both lay persons and (semi-)experts to which these texts are addressed. By contrast, factsheets emerged as the most "prototypically" popularized genre, positioning themselves halfway between specialized texts and the other popularized genres in terms of two out of three parameters of term density. The only category of terms for which factsheets are significantly denser than FAQs and news, and as dense as scientific opinions, is that of core terms: these, however, are systematically defined either in texts or within glossaries, possibly in an attempt to "[inform] and […] extend the reader's knowledge" (Gotti 2014, p. 17), the primary function of popularized texts.

To conclude, we would like to go back to the idea that vocabulary in popularized texts is chiefly "drawn from the general language" (Gotti 2011, p. 207; see Section 1). While the overall analysis of the vocabulary profiles of

EFSA texts provided support for the claim, the role of terminology in popularization discourse should not be underestimated: all of EFSA popularized texts were shown to display at least a degree of terminological density, and in this sense to display what Delavigne (2003, p. 84; my translation) calls the "discursive dilution of technical jargon". Somewhat paradoxically, popularized texts need to either explain or paraphrase technical terms in order to be understood by the audience; at the same time, however, they need these terms to gain legitimacy and authoritativeness in the eyes of the audience itself. Striking a balance between these two poles, as the analysis has hopefully suggested, is one of the biggest challenges for (successful) popularization.

Bionote: Adriano Ferraresi is Senior assistant professor at the Department of Interpreting and Translation of the University of Bologna, where he teaches a Master's module on information mining and terminology. His main research interests are in the area of quantitative, corpus-informed methods for the study of specialized varieties of English and in translation studies. He has given several invited seminars on these topics at various Italian and European universities. He recently published the monograph *Collocations in institutional academic English. Corpus and experimental perspectives* (Aracne).

Author's address: adriano.ferraresi@unibo.it

References

Andersen Ø. 1996, *Degree of specialization, network density and status - a text typological approach*, in Budin G. (ed.), *Multilingualism in specialist communication*, International Network for Terminology, Vienna, pp. 85-93.
Anthony L. 2014, *AntWordProfiler* [Software]. http://www.laurenceanthony.net/ (24.04.2019).
Baroni M., Bernardini S., Ferraresi A. and Zanchetta E. 2009, *The WaCky Wide Web: A collection of very large linguistically processed web-crawled corpora*, in "Language Resources and Evaluation" 43 [3], pp. 209-226.
Bondi M. 2015, *Language policy in web-mediated scientific knowledge dissemination: A case study of risk communication across genres and languages*, in Plo Alastrué R. and Pérez-Llantada C. (eds.), *English as a scientific and research language*, Vol. 2, De Gruyter, Berlin, pp. 85-111.
Cabré M.T. 2003, *Theories of terminology*, in "Terminology" 9 [2], pp. 163-199.
Cabré M.T. 2010, *Terminology and translation*, in Gambier Y. and van Doorslaer L. (eds.), *Handbook of Translation Studies*, Vol. 1, Benjamins, Amsterdam, pp. 356-365.
Calsamiglia H. and Van Dijk, T.A. 2004, *Popularization discourse and knowledge about the genome*, in "Discourse & Society" 15 [4], pp. 369-389.
Chung, T.M. and Nation, P. 2003, *Technical vocabulary in specialised texts*, in "Reading in a Foreign Language" 15 [2], pp. 103-116.
Constantinou M. 2014, *Conceptual metaphors of anger in popularized scientific texts. A contrastive (English/Greek/French) cognitive-discursive approach*, in Baider F.H. and Cislaru G. (eds.), *Linguistic approaches to emotions in context*, Benjamins, Amsterdam, pp. 159-188.
Corpas Pastor G. and Seghiri M. 2010, *Size matters: A quantitative approach to corpus representativeness*, in Rabadan R. (ed.), *Language, translation, reception. To honor Julio César Santoyo*, Universidad de Leon, Leon, pp. 111-145.
Coxhead A. 2000, *A new academic word list*, in "TESOL Quarterly" 34 [2], pp. 213-238.
Delavigne V. 2003, *Quand le terme entre en vulgarization*, in "Terminologie et Intelligence artificielle", pp. 80-91.
Freixa J. 2006, *Causes of denominative variation in terminology*, in "Terminology" 12 [1], pp. 51-77.
Gabrielatos C. and Marchi A. 2012, *Keyness: Appropriate metrics and practical issues*, paper presented at *Critical approaches to discourse studies 2012*, Bologna, 14 September 2012.
Gardner D. and Davies M. 2014, *A New Academic Vocabulary List*, in "Applied Linguistics" 35 [3], pp. 305-327.
Gotti M. 2011, *Investigating Specialized Discourse*, 3rd ed., Peter Lang, Bern.
Gotti M. 2014, *Reformulation and recontextualization in popularization discourse*, in "Ibérica" 27, pp. 15-34.
Ha A.Y.H. and Hyland K. 2017, *What is technicality? A Technicality Analysis Model for EAP vocabulary*, in "Journal of English for Academic Purposes" 28, pp. 35-49.
Krüger R. 2015, *The interface between scientific and technical translation studies and cognitive linguistics*, Frank &Timme, Berlin.
Kwary D.A. 2011, *A hybrid method for determining technical vocabulary*, in "System" 39, pp. 175-185.

Marín M.J. 2014. *Evaluation of five single-word term recognition methods on a legal English corpus*, in "Corpora" 9 [1], pp. 83-107.
Marín M.J. 2016, *Measuring the degree of specialisation of sub-technical legal terms through corpus comparison. A domain-independent method*, in "Terminology" 22 [1], pp. 80-102.
Myers G. 2003, *Discourse studies of scientific popularization: Questioning the boundaries*, in "Discourse Studies" 5, pp. 265-279.
Nation I.S.P. and Heatley A. 2002, *Range: A program for the analysis of vocabulary in texts* [Software]. http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx (24.04.2019).
Pearson J. 1998, *Terms in Context*, Benjamins, Amsterdam.
Piscopiello M.G. and Bertaccini F. 2009, *Densità e ricchezza terminologica: generi testuali a confronto*, in "MediAzioni" 7. http://mediazioni.sitlec.unibo.it (24.04.2019).
Rayson P. and Garside R. 2000, *Comparing corpora using frequency profiling*, in *Proceedings of Workshop on Comparing Corpora*, held at the ACL conference 2000, pp. 1-6.
Skibitska O. 2015. *The language of tourism: Translating terms in tourist texts*, in "Translation journal", October issue. http://translationjournal.net/October-2015/the-language-of-tourism-translating-terms-in-tourist-texts.html (24.04.2019).
Webb S. and Nation P. 2008, *Evaluating the vocabulary load of written text*, in "TESOLANZ Journal" 16, pp. 1-10.
West M. 1953, *A General Service List of English Words*, Longman, London.
Williams G. 1998, *Collocational networks: Interlocking patterns of lexis in a corpus of plant biology research articles*, in "International Journal of Corpus Linguistics" 3 [1], pp. 151-171.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 41-66
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p41

REPRESENTING AND RE-DEFINING EXPERT KNOWLEDGE FOR THE LAYMAN
Self-help medical manuals in late 19th century America*

GIULIANA ELENA GARZONE[1], PAOLA CATENACCIO[2]
[1]UNIVERSITÀ IULM, MILANO, [2]UNIVERSITÀ DEGLI STUDI DI MILANO

Abstract – This paper analyses a corpus (over 1 million words) of three self-help medical handbooks published in the US in the latter quarter of the 19th century, R.V. Pierce's *The People's Common Sense Medical Adviser* (1883), M.L. Byrn's *The Mystery of Medicine Explained* (1887), and *Gunn and Jordan's Newest Revised Physician* (1887). It aims to explore the discursive construction of medical knowledge and of the medical profession in the period, combining discourse analysis and corpus linguistics. The popularity of these manuals has to be seen within the context of medical care at a time when, in spite of the advances made in the course of the 19th century, the status of the medical profession was still unstable. Initially the focus of the study is on the representation of the medical profession. In this respect, the analysis testifies to an approach to traditional medical expertise which is essentially ambivalent, taking its distance from abstract medicine and quackery alike, while at the same time promoting a new approach based on different, more modern principles. The focus then shifts to the episteme of the medical science as represented in the works under investigation. The construction of selected epistemically relevant notions – *knowledge, theory/ies, experience, evidence,* and *observation* – is discussed relying on concordance lines in order to retrieve and examine all the contexts where they occur. The results of the analysis indicate a shift in the epistemological approach to knowledge, with theory and suppositions being complemented by experience, evidence and facts, and a representation of knowledge as a tool for empowerment, in line with the increasing democratisation of medicine characterising the period.

Keywords: medical knowledge; self-help medical handbooks; domestic medicine manuals; 19th century America; medical profession; democratisation of medicine.

* This study contributes to the national research programme 'Knowledge Dissemination across media in English: continuity and change in discourse strategies, ideologies, and epistemologies', financed by the Italian Ministry for the University (nr.2015TJ8ZAS).

1. Introduction

The focus of this paper is on self-help medical handbooks circulating in late 19th century America, a genre that was very popular, at a time and in a context where healing was still largely domestic in nature (Risse 1977, p. 1).

The health reform movement that had gathered momentum in the early half of the century had brought about a shift in the basic attitudes towards health, sickness and death. As Markell Morantz points out (1977, p. 79), sickness was no more tolerated with the stoicism of the colonials, and the idea emerged that health could be improved and disease at least partly prevented through individual effort. This, of course, enlarged the scope and intent of popularising medical handbooks.

This paper analyses a corpus of domestic medicine manuals published in the US in the latter quarter of the 19th century, and aims to explore and define the idea of medical knowledge underlying them. Medical knowledge is intended here not merely as pertaining to the medical sciences *strictu sensu*, their principles and practices, but also as encompassing various kinds of ancillary knowledge that may be described as “health education” – e.g. information and tips about personal and mental hygiene, notions about a healthy and active lifestyle, sanitary instructions, etc. –, which was also included in the handbooks themselves (McClary 1986).

1.1. Background: medicine and doctors in late 19th century America

Smith (1992, p. 251) relies on Slack’s (1979) statistics to point out that in the United Kingdom the publication of vernacular medical works and regimens started as far back as in the late 15th century, but gathered momentum in the second half of the 18th century, a growth that became even more marked in the course of the 19th century. At first, in the US domestic medicine manuals were imported from Britain. Cases in point are John Wesley’s *Primitive Physick*[1] (1747) and William Buchan’s *Domestic Medicine* (1769). In particular, the latter was the first in its kind for its completeness, containing also suggestions for prevention, and was translated into all major European languages. It enjoyed great popularity also in America, where various editions were published, enlarged and “revised and adapted to the diseases and climate of the United States” (under the editorship of Samuel Powel Griffitts, as specified on the title page of the 1795 edition). It maintained its popularity throughout the 19th century, with its last US edition in 1913 (Risse 1992, p. 186n.), and was taken explicitly as a model by later authors writing similar manuals (see e.g.

[1] The title was spelled *Physick* in the earlier editions and was changed to *Physic* in the 2oth edition in 1781.

Benezet 1826). For in the course of the century, in the US numerous domestic medicine manuals were written by practicing physicians, finding so much favour with the general public that they were reprinted in revised, updated and enlarged editions throughout the century and in some cases well into the 20th century.

The popularity of domestic medical manuals and health guides in 19th century USA has to be seen within the context of the situation of medical care at a time when the status of the medical profession was still unstable, especially in outlying rural areas where physicians were striving to establish the distinctive status it already enjoyed in Britain while lay healers practicing folk remedies were still very popular.

The lack of a solid reputation for the medical profession was partly due to the geographical characteristics of the country and the lack of a regulatory framework, notwithstanding the fact that since the late 18th century medical schools had started to be established (the first one in Philadelphia in 1765), medical societies founded (the Massachusetts Medical Society was incorporated in 1781) and exams for prospective doctors introduced in some States. Official recognition across the board, however, was still slow to come.

Another reason for the often doubtful prestige of the medical profession was that, although the 19th century saw dramatic advances and breakthroughs, with the gradual rise of the modern paradigm in medicine based on observation and experimentation (Furst 2000, pp. 4-5), in the early decades there was still a prevalence of traditional speculative medicine, based on abstractions (e.g. the theory of humours). Relying on this kind of "library" medicine, doctors were often uncertain and unsuccessful in their therapies, and this made it difficult for the layman to see in what respects they were better than quacks. For instance, the belief that diseases were caused by miasmas, and the lack of comprehension of contagion and sepsis often caused the catastrophic failure of many pharmacological and surgical therapies prescribed by physicians until late in the century (Furst 2000, p. 14). This undermined the reputation of the medical profession, which improved gradually as physicians became more knowledgeable and therefore better able to help their patients (Furst 2000, p. 16). In the meanwhile, until the second half of the century, in certain parts of the country it was admitted that anyone possessing some knowledge and/or some manual expertise (apothecaries, barbers, midwives, herbalists, clergymen, etc.; see Manger 2005, p. 301) could practice medicine. No wonder then that often people thought they could do without doctors, and opted for self-help, by "self-help" meaning "the diagnosis, care, and even prevention of disability and illness without direct professional medical assistance" (Risse 1977, p. 2).

According to Starr (1982), the tension between physicians' quest for a privileged status and popular resistance to their authoritativeness and power reflected "the conflict between a democratic culture and a stratified society"

(1982, p. 31). American society was growing increasingly democratic and egalitarian also thanks to the enlargement of the franchise, the spread of public education and the popular press. With the rise of people's confidence in their own competences and expertise, and of distrust in medical approaches which were not consistently successful, a cultural climate developed which was ideally suited to the spread of self-help manuals. The political climate of the mid-nineteenth century encouraged a shift from an almost mysteric approach to medicine to one based on "common sense": as Rutkow (2010, p. 45) points out, the "democratization of healing" – as she terms the shift – gathered momentum starting from "the ascent of President Andrew Jackson's populist embrace of the common man".[2] Indeed, "medical autonomy, non-interventionist care, and home doctoring were part of Jackson's call for self-determination in one's life" (Rutkow 2010, p. 45), and were also instrumental in fostering the rise of unorthodox, sectarian medicine, such as Thomsonianism (a botanically based movement),[3] Grahamism (based on a vegetarian dietary regime), hydropathy, and homeopathy (see also Hoolihan 2001, p. XV, and Duffy 1993, pp. 81-94), etc. All these movements were part and parcel of the transition (Haller 1981) which American medicine underwent in the second half of the 19th century.

Within this context, Starr (1982, p. 32) identifies three main spheres where medicine was practiced: domestic medicine, professional medicine practiced by physicians, and folk medicine practiced by lay healers. Among these, homecare was certainly the most popular and took advantage of healthcare manuals: so much so that often the self-help medicine of the period is referred to as "domestic healing" or "kitchen medicine" (Risse 1977, p. 3).

As Hoolihan puts it, "it is hard to overestimate the influence of popular medical literature as an instrument of reform" (2001, p. XV). While such popular literature was sometimes looked down on by some members of the medical profession, who also saw it as a form of competition, it was not infrequent for it to include serious scholarly efforts. It is meaningful that the most popular of these publications were written by physicians, especially by enlightened doctors who committed themselves to popularising medical knowledge and educating the general public. In time this contributed to spreading a scientific culture of medicine and respect for the medical profession's authoritativeness.

2 Jackson's Presidency: 1829-1837.

3 The denomination "Thomsonianism" derived from the name of this sectarian movement's founder, Samuel Thomson, whose book (Thomson 1825) was the basis of his botanico-medical movement.

2. Study design: aim, scope and materials

Against this backdrop, this paper investigates a corpus comprising three domestic medicine handbooks published in the US in the 1880s: R.V. Pierce's *The People's Common Sense Medical Adviser*, 1883 (11th edition), Marcus Lafayette Byrn's *The Mystery of Medicine Explained*, 1887, and *Gunn and Jordan's Newest Revised Physician*, 1887, presented as the 214th edition of John Gunn's famous book *Gunn's Domestic Medicine* (1832), revised and enlarged by Johnson H. Jordan, Gunn having died in 1863. [4]

Pierce's and Byrn's manuals had been published for the first time relatively recently, the former originally in 1875 and the latter in 1869 (copyrighted in 1887), with further revised and improved editions published in the following years, while Gunn and Jordan's *Newest Revised Physician* was the result of a historical accumulation of texts, as explained in the Preface. The original core of the 1,260-page tome was Gunn's volume *Gunn's Domestic Medicine* (604 pages), first published in 1832, which in time had had several "new and improved editions". In 1857 it appeared in a new edition entitled *Gunn's New Domestic Physician, or, Home Book of Health*: *A Complete Guide for Families* (1,129 pages). *Gunn and Jordan's Newest Revised Physician* was a further revised, updated and enlarged edition, copyrighted in 1885, and was defined as "the second family work originating with Doctor John C. Gunn" (1887: v). It featured a new version of the original manual radically revised, updated and enlarged by Dr Johnson H. Jordan, with a number of additions ("Anatomy", "Physiology and the Laws of Health", "The Vegetable Materia Medica," "Medical Recipes", etc.). It also included some further minor additions ("separate treatises on Anatomy, Physiology, and the Laws of Health") by Charles Alfred Rodin, M.D., the German co-translator of Gunn's original volume.

The history, origin and evolution of each manual are interesting in their own rights, and may be fruitfully investigated in order to trace the evolution of their contents over the decades; however, such an investigation is beyond the scope of this research, which aims to sketch a picture of the status of self-help books designed for lay users in the late 19th century. The three handbooks were chosen on account of their popularity during the 1880s (and, for that, for a long time after), as we wanted to focus our analytical efforts on texts that may be as representative as possible of the materials circulating among laymen in America at the time, to which ordinary people were exposed to in their search for information, advice or prescriptions regarding health, diseases, remedies and tips for a healthy lifestyle. Of course, these very popular manuals

4 The fac-simile editions of the books were retrieved from the U.S. National Library of Medicine Digital Collection (https://collections.nlm.nih.gov; accessed on 6 June 2017). To be analysed with computer routines, txt versions of the books were produced for machine readability.

contributed to constructing the idea of medicine and medical knowledge that prevailed in the period under examination.

Therefore, in this study special attention will be given to the function of these medical handbooks as instruments for the spread of medical knowledge among the lay population. Initially the focus will be on the representation of the medical profession; the focus will later shift to the episteme of the medical science as represented in the works under investigation, with special attention paid to the use of words connected with knowledge and its transmission and the discursive dynamics of which they are part.

Table 1 shows the details of the corpus:

	Gunn	**Byrn**	**Pierce**	**Overall corpus**
Number of words	616,021	210,498	322,325	**1,148,914**
Standardised TTR	42.12	43.78	45.09	**43.26**

Table 1
Corpus details.

As shown in Table 1, the corpus comprises more than a million words. The standardised TTR of the three texts ranges from 42.12 for Gunn to 45.09 for Pierce, averaging out at 43.26. Gunn's text appears to be the least lexically varied of the three – the larger size inevitably accounting for higher repetition rates of some lemmas.

The study takes a corpus based approach. This means that the bulk of the research is essentially qualitative, and computerised routines are relied on to calculate and compare the relative frequencies of selected lemmas across the three corpora and to retrieve contexts of usage, by making recourse respectively to the Wordlist and Concordance Tools of the Wordsmith Tools software suite (version 7; Scott, 2016). In other words, frequency data are not used to drive the investigation, but rather to corroborate hypotheses derived from close reading of representative parts of the texts, and in particular the introductions, and from contextual knowledge ensuing from an awareness of the socio-historical situatedness of the documents.

Because of the difference in the size of the various corpus components, wherever possible and appropriate frequency data are quoted in percentages. However, due to the low frequencies of some of the lemmas analyzed and the considerable differences in corpus size across the three corpora, normalisation of frequency data has often proved necessary, as suggested in various studies (see e.g. McEnery *et al*. 2006, pp. 52-53; McEnery and Hardie 2012, pp. 49-50), Data normalisation consists in calculating the frequencies that might be expected to be found in a corpus having the same characteristics as the one under investigation, but assuming its size to be an arbitrarily chosen number of words (typically 10,000, 100,000 or one million, depending on the original size of the corpus). Therefore, in the remainder of this essay in addition to frequency percentages, normalised (or relative) frequencies ("nf") are also given, setting the common base of normalization at 100,000 words. The formula used for the calculation is as follows:

nf = (number of occurrences ÷ number of tokens in the corpus) x (common base of normalisation = 100,000).

The analysis has been conducted on a set of lexical items deemed to be useful for the investigation of the construction of knowledge and the representation of episteme in the handbooks considered, making use of concordance lines to discuss their function in context.

Underlying this approach is the assumption that discursive artifacts both reflect and contribute to creating the conditions for social change (Fairclough 1992), of which the rise and vast circulation of medical self-help books was both a cause and a symptom. It has been mentioned in the introduction that the shift from traditional medicine to new approaches involved first and foremost a move away from abstract deductive reasoning, based on a priori unproven theories, to inductive, empirical methods of fact-finding in which evidence plays a key role. An essential component of this empirical turn was also the intent to explain the "mysteries" of medicine to the common folk, to quote from Byrn.

In the remainder of this essay the three texts under investigation will be explored to verify whether evidence can be found in them of an explicit textualisation of the shift in approach highlighted above.

3. Representation of the medical profession and self-help manuals' audience

As mentioned at the outset, all three manuals were written by medical doctors who took it upon themselves to instruct laypeople about the basics of healthcare. Representations of both authors and audiences are therefore embedded – explicitly or implicitly – in the texts of the manuals, and the exploration of such texts can help shed light on the ways in which both medical doctors and lay patients were constructed in the popularising medical discourse of the time. In particular, in respect of doctors, the ways in which they portrayed themselves as members of a disciplinary community, and the extent to which existing representations of medical professionals along the continuum from the quack to the scholar were drawn upon to establish their own professional and disciplinary identities (see Hyland 2012) may provide useful insights into the construction of the medical profession at the end of the 19th century.

In all the three handbooks analysed, the status of the authors as members of the medical profession is a crucial issue. On the front page of each volume, in addition to the author's name and academic title, quite interestingly short supplementary details are given to highlight the writer's authoritativeness:

> By M. Lafayette Byrn, M.D., Graduate of "The University of the City of New York," author of "Poisons in our Foods", etc, etc.
> By R.V. Pierce, M.D., founder of the Invalids' Hotel, and President for the World's Dispensary Medical Association

In the case of the revised edition of Gunn's handbook, no addition is made to the name of the original author (John C. Gunn,), but the qualification of the additional author is specified in terms of expertise:

> By J. H. Jordan, M. D., Physician to the Cincinnati Cholera Hospital in 1849.

In a similar vein, Byrn includes in his book a full transcript of his qualification as a M.D, explaining it as follows.

> there are so many imposters in large cities—men pretending to be physicians and surgeons, and have never been properly qualified to act in such capacity, I have concluded to produce the evidence for the information and benefit of those who do not know me personally, of my Medical Education having been regularly completed in one of the first Medical Schools in America. (Byrn 1887, p. 124)

The theme of impostors passing themselves off as qualified physicians recurs also in the other manuals. See for instance Pierce's comment of the system of "fake" medicine:

> I frequently receive letters making inquiries concerning the reputation or professional standing of a quack who resides in this city and makes a practice of imposing upon unfortunate sufferers. He assumes the title of "doctor" by virtue of a diploma obtained from a so-called medical university of Philadelphia, the charter of which has been revoked by the legislature of Pennsylvania, because the faculty of that institution were found guilty of selling diplomas to such charlatans. (Pierce 1883, p. 815)

The issue of the status of the medical profession and the competence of the books' authors as members of it are dealt with in the Introductions to the Manuals.

It is particularly meaningful that Dr. Gunn's Introduction published in the 1887 edition[5] of the *Newest Revised Physician* is integrally centred on the discussion of the status of the medical profession and the practice of medicine,

5 This Introduction first appeared in the 1963 edition of *Gunn's Domestic Medicine*, just before his death. That this Introduction was maintained unamended in the 1887 edition, and not only for documentary purposes (the text is not dated), shows that it was considered to be still topical and authoritative in the 1880s.

putting forth a very complex and dialectic argumentation. The main point made in the different stages of the discussion is the desirability of "a general spread of suitable knowledge among the people". Knowledge is presented as a form of empowerment, which, first of all, will defend individuals from "being made the easy prey of the villainous quack". This stance taken against lay healers represented as charlatans is not matched by a corresponding defense of physicians. On the contrary, within the medical profession two groups of doctors are declared unreliable: "eminent physicians who quarrel with each other for pre-eminence in fame, instead of endeavoring to enlighten and advance the happiness of the human family" and those who "disgrace their profession by sustaining the dark shadows of ancient superstitions, instead of advocating the improvements of modern times", and have "the rehearsal of former errors" as "the chief object of their works". This bipartite group of inadequate doctors is "dissociated" from the body of good and reliable physicians who "are well informed in their business", and possess "good judgement and common sense", which are founded on truth and experience, the best results often deriving from "simple remedies, and good nursing". Thus the rhetorical device of dissociation (Perelman, Olbrechts-Tyteca 1969, pp. 411ff.; van Rees 2007) is used to condemn some of the physicians, while avoiding the stigmatisation of the profession in its entirety.

In this respect, it is worth noticing that in the whole corpus the word *doctor*(*s*) has only 105 occurrences (just under 0.01%) although it is also used with a general meaning, which is often extended to refer to various kinds of healers (see below). This emerges clearly in the OED definition:

> Doctor 6.a.
> spec. A doctor of medicine; in popular current use, applied to any medical practitioner. Also: (amongst indigenous peoples) a traditional healer or diviner, esp. one dealing with afflictions thought to be caused by spirit possession or witchcraft. ("doctor, n." OED Online. Oxford University Press, June 2017. Accessed 25 November 2017.)

Another word that is used in the corpus, but less frequently, to refer to doctors is *practitioner*, which occurs 69 times (including the plural); this is its definition in the OED:

> Practitioner 1.a.
> A person engaged in the practice of medicine; a physician, surgeon, pharmacist, etc. ("practitioner, n." OED Online, Oxford University Press, June 2017, Accessed 25 November 2017.)

This is in contrast with the relatively high frequency of the word *physician*(*s*) which occurs 617 times (0.05%) and is by far the most frequent denomination for M.D.s in the corpus as a whole. This word more evidently refers to a qualified person who practices medicine, as in OED's definition:

Physician 1.a.
a. A person who is trained and qualified to practice medicine; *esp.* one who practices medicine as opposed to surgery. ("physician, n." OED Online, Oxford University Press, June 2017, accessed 25 November 2017.)

An examination of the concordance lines of the three words yields interesting indications as to their meaning and use at the time. It can be hypothesised that the word *doctor* is used sparingly because of its lack of a specific meaning: it appears in narratives of diseases and doctors, and to comment on doctors' bills, but it is also sometimes utilised in general terms to refer to those healers that are not considered to be legitimately practicing the profession, i.e. the "Botanic(al) doctors", the Uroscopian doctors who are described in a letter to the editor of the *Boston Medical and Surgical Journal* in 1845 as "a species of quack… who pretends to cure all diseases by examining the urine",[6] and also "Indian" doctors, whom Pierce (1883, p. 849) sees as "those having no knowledge of the delicate and intricate structure of the human system". But the word "doctors" is never used in those passages where medical practitioners are represented as authoritative sources of medical knowledge, but rather in more general contexts:

you only the price of one visit from a Doctor. Save expense by getting the "Family Doctor	Byrn
o do before you think of sending for a Doctor. It will make you a good nurse, and tell you	Byrn
t to do yourself. Everybody can't be a Doctor, but everbody can learn something about the	Byrn
unded with the illiterate "uroscopian" doctors, or fanatical enthusiasts, who ignorantly pret	Pierce
as favoring those itinerant self-styled doctors, or "professors," who perambulate the count	Pierce
h are compounded by quacks, "Indian doctors," or those having no knowledge of the delic	Pierce
together, to ask: " What shall he eat, Doctor?" "What can he eat?" "How often and when	Gunn
s the plant so much used by Botanical Doctors, called Thomsonians, supposed to have bee	Gunn
fashion, and late hours, the number of doctors, dentists, and apothecaries, and the amount	Gunn
ach. It has been justly said that, when doctors fail, the most simple remedies may effect a c	Gunn

Table 2
Selection of concordance lines for the lemma "doctor*".

This is in contrast with the use of the lexeme *physician*, which also sometimes appears in stories about sickness and medical care, but is regularly used to talk about physicians as depositories of medical knowledge and experience, and as sources of knowledge.

at each case must be investigated by the physician and treated according to its cause. For	Byrn
age the aid of an experienced and skillful physician who will interpose the required treat	Byrn

6 The letter, entitled "Medical practice and diseases in the West", was signed by A.B. Shipman, MD, Professor of Surgery in Laporte University, and published in the "Boston Medical and Surgical Journal" XXXII [7], 19 March 1845, p. 138.

d aether inhaled are used by surgeons and physicians, but are not safe in inexperienced ha	Byrn
pe of complete relief; but the experienced physician knows the treacherous symptom, and	Byrn
e Dr. Richard Bright, an eminent English physician, first recognized the affections and ma	Gunn
ions and careful watching of a competent physician can meet the necessities of such a case	Gunn
eneral terminated fatally. A distinguished physician of Moscow, in his report, states that dr	Gunn
and intemperance. Dr. Darbel, a French physician residing at Moscow, thinks that this at	Gunn
nts, or suffering from cold or exposure. A physician of Warsaw states in his report to the g	Gunn
is now admitted by the most experienced physicians in this disease, that this species of ul	Gunn
om its specific action upon serous tissues, physicians will readily understand how they can	Pierce
d not produce the desired result, a skillful physician's services should be secured, as he ma	Pierce
ections, often renders it difficult, even for physicians, to determine how far their treatment	Pierce
. This case had baffled the skill of several physicians, who regarded the disturbance of the	Pierce

Table 3
Selection of concordance lines for “physician”.

It emerges clearly that *physician* is the choice word to refer to the qualified and reliable medical professionals. It is used to represent the practitioner whose help is needed if one is to receive adequate treatment for serious or difficult to diagnose ailments, although criticisms are not spared, especially by Pierce (e.g. “it renders difficult, even for physicians to determine treatment shall be general and how far special”, “an error into which many physicians have fallen”: Pierce 1883, p. 497; etc.). But physicians are depicted as being the repositories of medical knowledge, whose views and opinions are to be relied on, so reference to their views counts as evidentiality. The opinions of eminent physicians, in many cases practicing in foreign countries (an eminent English physician, a physician of Moscow), are quoted as definitive sources of knowledge. It is noteworthy that among the collocates of *physician** there are (mostly in position L1) *skillful*, *competent*, *eminent*, *experienced*, *good*, *distinguished*.

In parallel, praise is repeatedly expressed for the ability of private individuals to contribute “information for the preservation of health and life of the most valuable character, solely derived from unstudied, or, at least, from unprofessional experience” (Gunn 1887, p. 7).

This position sanctioning the validity of lay experience, even when not supported by specialised study and professional training was already present in Buchan's popular domestic medicine manual, imported from Britain in the previous century, where it was stated openly:

> The knowledge of diseases does not depend so much upon scientific principles as many imagine. It is clearly the result of experience and observation.... Hence sensible nurses and other persons who wait upon the sick often know diseases better than those who have been bred to physic. (Buchan 3rd ed. 1769/1774, pp. 144)

Although Buchan himself assured that he did "not however mean to insinuate that a medical education is no use" (*ibidem*), his position was, as Lawrence (1975, p. 24) observes, "dangerously close" to extreme forms of anti-intellectualism, as well as a manifest "attack on an elitist status for the medical profession".

Also in this respect, in the above mentioned Introduction Gunn follows in Buchan's footsteps, when he asserts that medicine is not exclusively controlled by doctors, who do not have any monopoly:

> I am not attached to monopolies of any kind, and less than any to that which confines to a particular order that information which teaches how to relieve sickness and pain, and defend people's rights to apply their own remedies and to make recourse to physicians' assistance on a principled basis. (Gunn 1887, p. viii)

> It appears to me but fair to enlighten the people, as far as I can, on this important subject; for every one is interested in the prolongation of life and health, and should be, in a country like ours, allowed the privilege of thinking for himself, if he does not choose to act. (Gunn 1887, p. viii)

Thus, in spite of the progress that had been made in the course of the century, the diffidence towards professional medicine accompanied by a culture of self-help survived, although the stance taken by the author in this Introduction is probably influenced by old prejudices that younger practitioners had slowly been putting aside as the country moved on, but had not been completely eradicated.

In actual fact, the other two handbooks considered in this study do not deal with this issue in their Introduction, but – as will be seen – in other sections, as in the opening chapter they give preference to other aspects. The first one among such aspects is the theme of the democratisation of medicine. See the opening of Byrn's *The Mystery of Medicine Explained*:

> This book has been written for the "People!" the poor, the old, the young, male and female, the learned and the illiterate, those who are well and those who are sick; on land and on water, in the city and in the country, in the rural country-seat of the retired merchant or the log-cabin and camp-fire of the hardy pioneer or backwoodsman; for the clerk of sedentary habits, and for the farmer who toils in rain and in sunshine; for the young man far away from home, and for the mother who keeps watch over her loved ones through the long hours of dreary night in sickness;—in a word, for the million. (Byrn 1887, p. iii)

Not only does this opening convey a hyperbolic idea of the vastness of the potential audience of the book, but it also depicts it in terms of social variation, outlining an interesting picture of the American nation, including the well-to-do middle class but also pioneers and farmers – in a sort of thoreauvian picture of the group of intended addressees of the book.

The socially unlimited character of the potential audience for domestic medicine handbooks can also be found in the Prefaces to the other two works included in the corpus, but with a much lower profile. In the Dedication of Pierce's manual it is specified that the book is addressed to all of his patients, but the target audience seems to be even more all inclusive in geographical terms, virtually without boundaries

> To my patients who have solicited my professional services from their homes, in every state, city, town, and almost every hamlet, within the American Union, also to those dwellings in Europe, Mexico, South America, the East and West Indies, and other foreign lands I respectfully dedicate this work. (Pierce 1983, p. iii)

This obviously suggests that the author sees his book's potential audience as planetary, and at the same time implies that the writer's reputation as a physician is universal.

Strictly connected with this is in all three books the promise to use "simplest language, adapted expressly to the use of families" (Gunn 1887, p. viii), "information couched in language free from medical technicalities" (Byrn 1887, p. iii), with each author proposing "to express himself in plain and simplest language, and, so far as possible, avoid the employment of technical words" (Pierce 1883, p. vi). Such insistence on the use of plain language can be seen – at least in part – as indebted to the culture of accessibility and clarity in scientific writing originally introduced and promoted by the founding members of the Royal Society in the course of the 17th century (see, amongst others, Moessner 2009). However, the fact that all authors insist that a lay audience should be able to understand the handbook testifies to a strong democratising thrust which is much more comprehensive in nature and outlook.

A common denominator is the idea of knowledge as a form of empowerment. If, as seen above, the status of the medical profession is still to some extent questioned, all three authors assert the need for the layman to possess enough knowledge to "be enabled to give the best remedies, where a physician cannot be had, or, in cases of emergency, to know what to do before the physician arrives, so as to alleviate suffering or be the means of saving life." (Byrn 1887, p. iii). In the case of Pierce's book, the acquisition of knowledge is also seen as a form of empowerment to avoid problems deriving from inappropriate marriage (based on the Theory of Temperaments!) and to prevent "unfortunate young men and women" from having to suffer the consequence "of certain abuses, usually committed through ignorance" (Pierce 1883, p. vi) , presumably venereal diseases and unwanted pregnancies.

Thus, access to knowledge is presented as one of the most effective tools to prevent and manage health problems and disease.

4. Defining medical knowledge: issues of access and co-construction

In light of the above, it is to be expected that the authors' presentation of their remedies and recommendations should not only be couched in accessible language, but also framed within a discourse of knowledge construction which, on the one hand, takes its distance from abstract medicine and quackery alike (see Section 1.1 above), and, on the other, promotes a new approach based on different principles. Indeed, the discursive construction of expert knowledge is central to the manuals, and its investigation can provide useful insights into the changing episteme of medical knowledge. In order to explore this topic, the construction of certain epistemically relevant notions is discussed, relying on concordance lines for the relevant words: *knowledge, theory/ies, experience, evidence*, and *observation*.

Attention is first addressed to the word *knowledge*. The linguistic representation of the notion it refers to is the main object of this discussion given that it can be seen as a superordinate of the other lexemes to be analysed, all of which fall within its overarching semantic area, so that the examination of the latter can be useful to further outline and specify what counts as knowledge in these popular medicine handbooks, and how such knowledge is textually constructed.

The lemma *knowledge* has noteworthy frequencies in Gunn's and Pierce's books, while it is far from frequent in Byrn's:

	overall	Gunn	Byrn	Pierce
knowledge	150 (0.01%)	78 (0.01%)	7 (<0.01%)	65 (0.02%)
normalised frequency	13.05	24.34	3.32	20.16

Table 4
Frequency of the lemma *knowledge*.

It can be noted that the frequency found is much lower in Byrn's book than in the other two. As will be seen, this is a constant element emerging from the analysis of all the different lemmas having epistemic significance, possibly indicating a less speculative attitude on the part of an author whose attention tends to be focused prevalently on practical and contingent aspects. But apart from quantitative divergences in the three authors' choices, within the corpus as a whole the notion of *knowledge* is generally presented as the guiding principle that makes it possible to understand material facts rationally, e.g.:

> It is our knowledge of organic chemistry which is guiding us to a rational comprehension of the utility of food and the requirements of the organism. (Gunn 1887, p. 34)

Conversely, lack of knowledge may have dramatic consequences:

> From the lack of this knowledge on the part of parents, many a little one has perished before medical assistance could be obtained (Byrn 1887, p. 128)

That knowledge provides empowerment is not only a recurring theme in these manuals, but it is also stated openly in Pierce's manual:

> "Knowledge is Power." That knowledge which is conducive to self-preservation is of primary importance. […] Believing that the diffusion of knowledge for the prevention of disease is quite as noble a work as the alleviation of physical suffering by medical skill, I have devoted a large portion of this volume to the subjects of physiology and hygiene. (Pierce 1883, p. 388)

It is argued here that knowledge is most effective in preventing disease, presenting its dissemination not so much as a way to further intellectual progress, but as functional to its practical application. This somehow explains why in handbooks that purport to be aimed at self-help and practical applications there are so many parts that are entirely devoted to explaining scientific concepts. However, knowledge is never imparted as an end in itself or merely as a tool for intellectual improvement, but as a powerful instrument to educate the general public to preserve their health and help them to learn from experience, doing away with antiquated and ineffective forms of deductive reasoning based on abstract principles seldom anchored in experience and observation. Pierce states this clearly:

> But knowledge is being diffused, education is lifting the masses, and dear-bought experience is opening the eyes of thousands, who now believe in hygiene and remedial restoration, rather than in the employment of debilitating, exhausting and disease-creating medicines. (Pierce 1883, p. 570)

This quotation from Pierce's book summarises very well the paradigm shift in the construction of knowledge from an idea of the medical science as based on abstractions and unquestioned conceptualisations to be applied by default, indiscriminately, to contingent cases, to one based on a more rigorous scientific approach. This change in perspective entails first a loosening of previous restrictions and prejudices towards laypeople's access to knowledge and, secondly, a broadening of the principles upon which such knowledge is co-constructed through different types of experience, including people's "common sense".

Among the lemmas referring to notions that are hypo-ordinates to *knowledge*, the first lemma to be discussed here is *theory* (in both its singular and plural form, *theor**), considering that traditional medicine is represented in the manuals as being essentially based on a priori speculations and theories (e.g. miasma theory). Quite interestingly, this lemma has a rather low frequency in the corpus and only occurs in two of the three texts, being pointedly absent from Byrn's. Pierce seems to resort to the word *theory* the most, with almost twice as many occurrences as in Gunn's book, but in a text half the size, as highlighted by the normalised frequency figures.

	overall	Gunn	Byrn	Pierce
theor*	97 (<0.01%)	25 (<0.01%)	0	42 (<0.01%)
normalised frequency	17.08	4.05	0	13.03

Table 5
Frequency of the lemma *theor**.

While not numerically salient in themselves, occurrences of *theor** do appear to play a pivotal role, argumentatively, in the two handbooks in which the lemma occurs, as the examples below illustrate.

In most cases the word is used to mean "Abstract knowledge or principles, as opposed to practical experience or activity; theorising, theoretical speculation", rather than "The conceptual basis of a subject or area of study (see OED Online 2017: "theory, n.", 1 and 3); in other words, it tends to be used to refer to speculation rather than to establish scientific knowledge. See the following example from Gunn (1887):

> Treatment of the Cattle Plague or Rinderpest. — 1. Vaccination. This has been recommended, on the theory that the Cattle Plague is analogous to or identical with Small-pox. Vaccination with the lymph of Cow-pock may then be resorted to as a prophylactic. It may be practiced on the udder or vulva of a cow. It has already been resorted to very extensively, but with doubtful success. (Gunn 1887, p. 40)

In some cases, 'theories' are contrasted with 'facts of science':

> My object is to inculcate the facts of science rather than the theories of philosophy. (Pierce 1883, p. 8)

In the following example, commenting on the medical profession's resistance to new ideas, theories become worthy of respect only when they are proved by facts:

> Thus has it ever defended its established opinions against innovation; yet, out of this very conservatism has grown much real good, for, although it has wasted no time or energy in the investigation of theories, it has accepted them when established as facts. In this manner it has added to its fund of knowledge only those truths which are of real and intrinsic value. (Pierce 1883, p. 294)

Sometimes, as in the example below, theory – in this case encoded by means of a verb form – is explicitly contrasted with practice, with the latter getting the upper hand:

> Men may theorize finely, but at the bedside practice unsuccessfully: in preference to such persons, give me a good old woman, with her teas and pimples, and I will trust the rest to nature. (Gunn 1887, p. 25)

On the opposite end of the theory-empiricism scale are words such as *experience*, *evidence*, and *observation* which highlight the empirical, experimental and observational status of medical knowledge.

The next lemma to be analysed here is *experience*, which represents an important pole in the epistemic view presented in the manuals under discussion, as it embodies the tendency to leave abstractions behind and give preference to a more empirically based approach. The relevant frequencies are shown in Table 6.

	overall	Gunn	Byrn	Pierce
experience	185 (0.02%)	92 (0.01%)	14 (<0.01%)	79 (<0.01%)
normalised frequency	16.18	8.00	1.21	6.87

Table 6
Frequency of *experience* (noun).

If one compares these data with those in Table 4 regarding *knowledge*, it emerges that the overall frequency of the lemma *experience* is considerably lower than that of *knowledge*. This could lead to the hypothesis that *experience* occupies a lower epistemic status than *knowledge* in the texts at hand. However, as shall be seen, *experience* is one of the means through which knowledge can be obtained – others being, for example, *observation* and *evidence*. It would appear that *experience* is one of the means to obtain knowledge, though by no means the only one. It should be noted that also in this case Byrn's text features considerably lower frequencies than the other two authors, as it does for *knowledge* and *theor**, which would appear to support the hypothesis put forth above that in his book Byrn is less concerned than his colleagues with reflecting on the practical knowledge he imparts and on how it originated.

The collocates are also interesting, as in most cases they highlight the fact that experience is represented as one of the sources of knowledge which – if intense and repeated – is noteworthy and reliable, as the collocates of the word show. They are: *extensive* 5 hits (left), *large* 2 hits (18 left, 2 right), *long* 18 hits (16 left, 2 right), *proved* 12 hits (2 left, 10 right), *great* 11 (4 left, 7 right), all adjectives qualifying the kind of experience needed, and *observation* 9 hits (5 left, 4 right), a noun which often appears in the pattern "*experience and observation*". This is because experience is occasionally represented in combination with observation and also with "unfettered intelligence", as useful to prove or demonstrate certain facts and truths on the basis of which physicians can proceed to diagnose and treat diseases.

This is what emerges from a perusal of a selection of representative concordance lines for *experience** (noun):

ifficult to remove. It has been found by experience, that rubbing the child over with hog's l	Byrn
valuable guides they are at times, when experience, observation, etc., have given the powe	Byrn
the infant. Nature, therefore, as well as experience, indicates the propriety of with holding	Byrn
acquired only by attentive practice and experience, aided by previous anatomical knowled	Byrn
e, and Remedy. Strange as it may seem, experience has proved that, after great fatigue, the	Gunn
cessity for it whatever. Discoveries and experience of late years have amply demonstrate	Gunn
s, averted and cured diseases that bitter experience tells us have proved fatal for want of o	Gunn
urious. I have long been satisfied, from experience and observation, that much of the coug	Gunn
sunderstood. They speak the feelings of experience, of unfettered reason and observation,	Gunn
urated with chemical properties, which, experience has taught us, exercise a specific antise	Pierce
dropsy. The specialist, skilled by large experience in detecting the exact morbid condition	Pierce
country who have not, by observation or experience, become somewhat familiar with this	Pierce
may be forwarded to a specialist of large experience in this disease, who will readily deter	Pierce

Table 7
Selection of concordance lines for "experience".

The importance of experience is confirmed by a further 156 occurrences of the cognate verb *to experience* and 40 hits for the relevant adjective (*experienced*).

Some examples of the use of *experience* and cognate words will be useful to illustrate the contexts where they are used, mostly with reference to physicians or other health professionals, explicitly or metaphorically: *competent and experienced physician* (Byrn), *skilled and experienced physician* (Pierce), *careful and experienced accoucher, or midwife*, *practical and experienced physician*, *steady and experienced hands* (Gunn).

In some cases, the command of a good record of experience is presented as one of the requisites of the reliable physician:

> It is the imperative duty of every sufferer from this disease, no matter how seldom the unnatural losses occur, to engage the services of a competent and experienced physician (Byrn 1887, p. 258)

More often e*xperience* indicates a fact from which something is learnt, being followed by verbs like *to show, to confirm, to demonstrate*, e.g.:

> Nurses, doubtless, are sometimes found to whom a child may be safely intrusted [sic]; but experience has but too often shown that the reverse is the case.
> (Byrn 1887, p. 80)

> This dose of Oil may seem large for a child of that age, but experience has confirmed the safety and great benefit derived in this complaint from large doses of Oil. (Gunn 1887, p. 226)

> It has been found by experience, that rubbing the child over with hog's lard, until it becomes completely incorporated and mixed with this substance, and then making use of soft dry flannel to remove it, is the most simple, expeditious and perfect manner of getting rid of it. (Byrn 1887, p. 104)

Here *experience* is constructed as an essential source of knowledge, as it overrides common sense and principles, also offering a chance of serendipity in finding therapeutic solutions.

The analysis of the lemma *experience* and its partial overlap with or, more aptly, subordination to *knowledge* (in so far as experience leads to knowledge) discussed in the previous paragraphs confirms the heightened value placed on empirical sources of knowledge with respect to traditional approaches. This divergence takes the form – amongst other things – of a shift from an epistemological approach to knowledge based on *theory* to one grounded in *experience*, *evidence* and *observation*. The latter two are therefore the words whose use in the texts will now be considered.

The lemma *evidence* is very unevenly distributed across the three texts. As with *theor**, Gunn and Pierce feature it the most, while Byrn has the lowest number of occurrences.

	Overall	Gunn	Byrn	Pierce
evidence	116 (<0.01%)	44 (<0.01%)	12 (<0.01%)	50 (<0.01%)
normalised frequency	28.35	7.14	5.7	15.51

Table 8
Frequency of the lemma *evidence.*

Here again differences are noticeable between Gunn and Pierce on the one hand, and Byrn on the other.

In Gunn and Pierce *evidence* is used either to drive home the reliability of scientific claims made in the text, or to deny the reliability of commonly held beliefs:

> Lightning, or electricity, obeys one unvarying law. It uniformly follows the best continuous conductor, and no conductor can be considered good unless it is continuous. Abundant evidence of this is afforded by the use of broken or otherwise defective lightning rods. (Gunn 1887, p. 1077)

> There is no satisfactory evidence that the fruits and condiments brought to us from the tropics have any other than a pleasurable and beneficial effect, when used with reasonable caution and with a due regard to individual idiosyncrasies. (Gunn 1887, p. 781)

> Tomatoes were formerly regarded as only fit for hogs, and their introduction to fashionable tables in this country in raw and stewed form is of very recent

> date. Their value as medicinal food has probably been much overestimated, and is based chiefly, I think, on the same theory that has given Graham bread such great prominence, namely, that the value of any article of food is in pretty exact proportion to its disagreeableness, since, to the majority of persons, tomatoes are at first repulsive. There is no evidence that they are in any respect superior to ripe peaches, or pears, or apples. (Pierce 1883, p. 215)

In the latter example, Graham's theory (a popular one at the time of publication of the manual) is evoked only to be dismissed because of *lack* of evidence. Thus, *evidence* appears to be used to disprove wrong beliefs, as well as to put forth facts or scientifically reliable truths.

By contrast, examples of the use of *evidence* in Byrn's manual do not seem to fall along the same line as in the passages quoted above. Rather, Byrn hardly ever uses *evidence* to contrast it with previously held beliefs, but rather as an element emerging from observation that may be used as a starting point for trying a therapy, as can be seen in the following example:

> The power of belladonna, in protecting individuals against the contagion of scarlet fever, has been much discussed. It has been used extensively, and with apparent success; at all events, the evidence is sufficient to make it worth a trial during the prevalence of a very severe or malignant form of scarlet fever. (Byrn 1887, p. 129)
>
> The condition and appearance of the tongue, are indications almost always consulted by a physician in investigating a case of disease, and most valuable guides they are at times, when experience, observation, etc., have given the power of reading them aright. When the appearances of the tongue, however, are admitted as evidence, consideration must always be given to the natural state of the organ in the individual … (Byrn 1887, p. 425)

Thus, while Byrn does occasionally use *evidence*, the discursive function of the term does not appear to be as salient, as it seems to be prevalently used as a synonym for 'realisation' with an essentially 'local' meaning, while in the other two texts the word is deployed in reasoning in the service of the epistemic evaluation of facts observed or data gathered.

In Gunn's and Pierce's books the programmatic use of *evidence* in aid of experimental and/or empirical science is emphasised by its attributive patterns. The lemma is typically accompanied by adjectives such as *conclusive, undisputable*, *positive*, *unmistakable, abundant, unequivocal* and *incontrovertible* in attributive position. The use of these accompanying adjectives reinforces the epistemic strength of the statements in which they occur.

en, bilious, sour, bloody or black matter, is evidence of disease of the brain, stomach, live	Byrn
nd with apparent success ; at all events, the evidence is sufficient to make it worth a trial d	Byrn
e to a river, that is thought to be conclusive evidence that he is not mad. But the truth is, th	Byrn
eatment. Still we have, after all, conclusive evidence to show that sensible changes and ext	Gunn

swallowed about that time, is unequivocal evidence of its importance to the digestive orga	Gunn
hich is another pretty reliable corroborative evidence. There are various other evidences of	Gunn
nd is often referred to as a most remarkable evidence of design on the part of the Creator.	Gunn
not to be accepted, however, as conclusive evidence of the existence of stone in the bladd	Pierce
nsibilities sufficiently to give unmistakable evidence of his masculine attributes. One boy	Pierce
readers for here offering some indisputable evidence of the extraordinary success which I	Pierce
wonder to all who know me, and a positive evidence of your Superior skill and potent rem	Pierce

Table 9
Selection of concordance lines for *evidence*.

The consistent use of a premodifier to accompany *evidence* seems to imply that "evidence" in itself is not sufficiently meaningful to be taken into considerations, while it is its conclusive, positive, powerful, remarkable, satisfactory or real character that makes it meaningful. Similar comments apply to the lemma *observation*, which has frequencies that are very similar to those of *evidence*:

	Overall	Gunn	Byrn	Pierce
observation*	105 (>0.01%)	45	10	50 (0.01%)
normalised frequency	27.56	7.3	4.75	15.51

Table 10
Frequency of *observation**.

Also in the case of *observation*, there are slight differences in usage among the three authors, as emerges from a selection of the concordance lines:

g of an infant thus circumstanced. The same observations apply to infants, whose stomach	Byrn
g in any or every posture, are all matters for observation; also any habitual cough, and its c	Byrn
illness and oppression, which soon goes off. Observation and facts show, however, that the	Gunn
. Often there is a great deal of reason in the observation. Much of the bile secreted by the l	Gunn
ems to be acting vicariously. Only the close observations and careful watching of a compe	Gunn
ings of experience, of unfettered reason and observation, Which fully confirm that solemn	Gunn
yed cells. The intellect, whether engaged in observation, generalization or in recondite stu	Pierce
brain, but the others, as is clearly shown by observation and experiment, cannot be restrict	Pierce
hey had perforated the intestine; but careful observations have proved that they can only e	Pierce
ars, has been considerably reduced. Clinical observation proves that injuries to the lungs a	Pierce

Table 11
Concordance lines for *observation.*

An examination of the concordance lines shows a difference between Gunn and Pierce on the one hand, and Byrn on the other. Gunn and Pierce tend to use this word to refer to the action of observing scientifically, often in combination with words like *experience* and *investigation*, and premodified by adjectives like *clinical*, *close*, and *careful;* Byrn, on the other hand, uses it less frequently, in half of the cases in the plural and more often with a contingent meaning referring to rules or generalisations gathered from experience. See the following examples are given below:

> But the cause can generally be ascertained by careful observation; then we can generally remove the symptom by removing the cause […] (Gunn 1987, p. 320)
>
> There is, perhaps, no subject more interesting or important for investigation and observation, than the Diseases of Children […] (Gunn 1987, p. 589)
>
> Clinical observation proves that injuries to the lungs are not so fatal as was once supposed. (Pierce 1883, p. 496)
>
> It should be expected that such careful, or pains-taking experiments, as are necessary to establish a science, will be preceded by intuitive judgment and accredited observations, which may be, for a time, the substitutes of those more abstruse in detail. (Pierce 1883, p. 140)
>
> In consequence of the deterioration of the mind which the disease occasions, development of the mental functions are sadly interfered with, capacity of acquirement is lessened, progress is arrested, and hence the frequent observation of the precocious youth becoming the dull adult. (Byrn 1887, p. 255)
>
> While making these observations, however, it must not be presumed that we ignore the culture of the mind or of the taste (Byrn 1887, p. 493)

Although – as mentioned above – Byrn tends to use the word *observation* with a more contingent meaning, this does not diverge from an approach oriented towards empirical observation and the gathering of evidential proofs rather than to the application of abstract and unproven principles.

Overall, also the analysis of the usage of the lexeme *observation* confirms a conceptualisation that is part of a method leaving behind unproven theories based on abstractions in favour of an inductive, fact-based approach in the medical sciences, as explicitly asserted in Pierce's statement above (the fourth in the sequence of examples above), advocating that intuitive judgment and observation, followed by careful or painstaking experiments, may replace approaches that are "more abstruse in detail".

The analysis conducted identifies some common traits across the three textbooks, but also highlights some diverging trends. In particular, Gunn and Pierce appear to share a more speculative focus – though consistently within the boundaries of evidence-based medicine – whereby they constantly strive to instruct the general public on the basics of medical self-help, but also to foster forms of critical appraisal of medical knowledge. Byrn, on the other hand, seems to eschew all forms of speculation, preferring to target more practical aspects. His approach is one based on practical knowledge, with little attention paid to sources of knowledge, and with the prevalent aim to provide practical advice to be relied upon in the absence of medical help.

5. Conclusions

The analysis conducted on three medical manuals written for the benefit of lay people suggests that the authors of these publications were very well aware of their books' usefulness in helping people prevent health problems and solve them when they arose, except for the most serious conditions. The introductions to the texts confirm that while an eminently practical intent was a key determinant of the rise of such publications, conventional medical expertise being often hard to come by especially in isolated areas, or under circumstances in which getting hold of a doctor was not expeditious (for example in outlying rural areas or wild parts of the country, or at sea), a polemical intent was also prominent, with dissatisfaction with the approach of traditional medicine playing an important role in the framing of medical knowledge.

The texts testify to an ambivalent approach to traditional medical expertise: while doctors are still represented as playing a key role in the transmission of medical knowledge, a distinction is drawn between good and bad doctors, the latter being depicted either as incompetent charlatans or – at the other extreme – as physicians possessing a purely theoretical expertise, based on more or less traditional or sectarian unproven principles, capable only of giving advice that does not pass the test of empirical evidence.

The texts also testify to a shift in the epistemological approach to knowledge, with theory and suppositions being complemented and supported by experience, evidence, and observed facts. In this respect, two of the texts – Gunn's and Pierce's – appear to be more polemical in outlook, with theory being mentioned in mostly derogatory terms and frequently contrasted with factual evidence. This seems to be a remainder of the anti-intellectualism and empiricism about the medical profession prevailing in the previous century, which was hard dying in spite of the great progress of the medical sciences in the 19th century. In the case of Gunn's book this may be due to its being based on a work originally written in 1832, in spite of the radical revisions, updates, enlargements and additions made in the following decades, but it is certainly first-hand in the case of Pierce's volume which first appeared in the early 1880s. The same stance also features in Byrn's book, but the contrast with theory is not topicalised as it is in the other two texts, as this manual tends to be more exclusively practical in approach, with limited scope for reflection.

It is interesting that, in spite of the emphasis placed on experience and evidence, all three handbooks devote entire and lengthy chapters to imparting their readers general knowledge not only of the functioning of the human body, but also of more general topics in Anatomy, Physiology, Botanics, Domestic and Sanitary Economy, Emotions, and Life and Morals.

The significance of this comprehensive approach becomes clear if one considers that a common denominator of the manuals is the representation of knowledge as a tool for empowerment, in line with the increasing democratisation of medicine characterising the period, also praising lay common sense for providing useful insights which occasionally contradict, but more often supplement medical knowledge. This suggests an important step towards the co-construction of knowledge between expert and layman – a crucial aspect of the democratisation of science.

Bionotes: Giuliana Garzone is Full Professor of English, Linguistics and Translation at IULM International University of Languages and Media, Milan, where she directs the MA in Specialised Translation and Conference Interpreting. Previously she taught at the University of Milan and the University of Bologna. Her research interests are mainly in ESP, which she has explored in a discourse analytical perspective, integrating it with corpus linguistics. She has co-ordinated several research projects and published extensively on legal, scientific and business discourse as well as on translation and interpreting. She is editor-in-chief of the journal *Languages Cultures Mediation* and of the associated book series, and is co-editor of the series "Lingua, traduzione, didattica" (FrancoAngeli). She sits on the Advisory Board (or Scientific Committee) of the journals *Text & Talk*, *Journal of Multicultural Discourses*, *Lingue e Linguaggi* and *Euroamerican Journal of Applied Linguistics and Languages.*
Paola Catenaccio is Full Professor of English Linguistics and Translation at Università degli Studi di Milano. Her research interests lie primarily in the field of discourse analysis, which she applies to a variety of domains (legal discourse, business communication, professional discourse, ELF communication, the discourse of science and of scientific popularisation) in combination with other methodological perspectives (most notably corpus linguistics), adopting a multi-methods approach to linguistic research. She has authored numerous articles which have appeared in international journals and edited collections. She has also coedited volumes on various aspects and domain-specific discourse (*Genre Change in the Contemporary World*, 2012, with G. Garzone and C. Degano; *Professional Practice across Domains: Linguistic and Discursive Perspectives*, 2018, special issue of the Journal *Languages Cultures Mediation*, with G. Garzone and S. Sarangi), and authored volumes on the interface between corporate communication and the media (corporate *Press Releases. An Overview*, 2008) and on emerging paradigms in business discourse (*Understanding CSR Discourse: Insights from Linguistics and Discourse Analysis*, 2012).

Authors' addresses: paola.catenaccio@unimi.it, giuliana.garzone@iulm.it

References

Duffy J. 1993, *From Humors to Medical Science. A History of American Medicine*, University of Illinois Press, Urbana and Chicago.

Fairclough N. 1992, *Discourse and Social Change,* Polity Press, Cambridge.

Furst L.R. (ed.) 2000, *Medical Progress and Social Reality. A Reader in Nineteenth Century Medicine and Literature*, State University of New York Press, Albany.

Haller J.S. 1981, *American Medicine in Transition 1840-1910*, University of Illinois Press, Chicago.

Hoolihan C. 2001, *An Annotated Catalogue of the Edward C. Atwater Collection of American Popular Medicine and Health Reform*, The University of Rochester Press, Rochester.

Hyand K. 2012, *Disciplinary Identities. Individuality and Community in Academic Discourse*, Cambridge University Press, Cambridge.

Lawrence C.J. 1975, *William Buchan: Medicine Laid Open,* in "Medical History" 19, pp. 20-35.

Markell Morantz R. 1977, *Nineteenth Century health Reform and Women: A Programme of Self-Help*, in Risse G.B., Numbers R.L. and Walzer Leavitt J. (eds), *Medicine without Doctors*, Science History Publications, New York, pp.73-93.

Manger L.N. 2005, *A History of Medicine*, Taylor & Francis, Boca Raton.

McClary B.H. 1986, *Introducing a Classic: "Gunn's Domestic Medicine",* in "Tennessee Historical Quarterly" 45 [3], pp. 210-216.

McEnery T. and Hardie A. 2012, *Corpus Linguistics: Method, Theory and Practice*, Cambridge University Press, New York.

McEnery T., Xiao R. and Tono Y. 2006, *Corpus-Based Language Studies. An Advanced Resource Book*, Routledge, London.

Moessner L. 2009, *The Influence of the Royal Society on 17th-Century Scientific Writing*, in "ICAME Journal" 33, pp. 65-87.

Perelman C. and Olbrechts-Tyteca L. 1969, *The New Rhetoric: A Treatise on Argumentation* (*Traité de l'argumentation - la nouvelle rhétorique*, 1958 translated by J. Wilkinson and P. Weaver), Notre Dame, University of Notre Dame Press.

Porter R. (ed.) 1992, *The Popularization of Medicine 1650-1850*, Routledge, London.

Risse G.B. 1977, *Introduction*, in Risse G.B., Numbers R.L. and Walzer Leavitt J. (eds), *Medicine without Doctors*, Science History Publications, New York, pp. 1-9.

Risse G.B. 1992, *Medicine in the Age of the Enlightenment*, in Wear A. (ed.) *Medicine in Society: Historical Essays*, Cambridge University Press, Cambridge, pp. 149-195.

Rutkow I. 2010, *Seeking the Cure. A History of Medicine in America*, Scribner, New York.

Scott M., 2016, *WordSmith Tools version 7*, Stroud, Lexical Analysis Software.

Slack P. 1979, *Mirrors of health and treasures of poor men; the use of vernacular medical literature in Tudor England*, in Webster C. (ed.) *Health, Medicine and Mortality in the Sixteenth Century*, Cambridge University Press, Cambridge, pp. 237-273.

Smith G. 1992, *Prescribing the rules of health: self-help and advice in the late eighteenth century*, in Porter R. (ed.) *Patients and Practitioners. Lay Perceptions of Medicine in Pre-Industrial Society*, Cambridge University Press, Cambridge, pp. 249-282.

Starr P. 1982, *The Social Transformation of American Medicine*, Basic Books, New York.

van Rees A. 2007, *Dissociation in Argumentative Discussions. A Pragma-Dialectical Perspective*, Springer, Doerdrecht.

Worboys M. 2011, *Practice and the Science of Medicine in the Nineteenth Century*, in "Isis. A Journal of the History of Science Society" 102 [1] (March), pp. 109-115.

Self-help Manuals Cited[7]

Benezet A.A. 1826, *The Family Physician., or, a Treatise on the Prevention and Cure of Disease, Calculated Particularly for the Inhabitants of the Western Country, and for Those Who Navigate Its Waters*, W.H. Woodward, Cincinnati.

Buchan W. 1769, *Domestic Medicine, or the Family Physician*, Balfour, Auld and Smellie, Edinburgh.

Buchan W. 1795, *Domestic Medicine, or a Treatise on the Prevention and Cure of Diseases, by Regimen and Simple Medicines*, revised and adapted by S. Powel Griffitts, T. Dobson, Philadelphia

*Byrn M.L. 1887, *The Mystery of Medicine Explained. A Family Physician and Household Companion, Prepared for the Use of Families, Plantations, Ships, Travelers, & C.*, 81st ed., Coast City Publishing Company, New York.

Gunn J.C. 1832, *Gunn's Domestic Medicine: or Poor Man's Friend*, F.S. Heiskell, Knoxville, T.

Gunn J.C. 1857, *Gunn's New Domestic Physician, or, Home Book of Health*. *A Complete Guide for Families*, Moore, Wilstach, Keys & Co., Cincinnati.

*Gunn J.C. and Jordan H.J. 1887, *Gunn and Jordan's Newest Revised Physician: Being the First New Domestic Physician or Home Book of Health*, 214th ed., J.E. Moore & Company Publishers, Cincinnati.

*Pierce R.V. 1880/1883, *The People's Common Sense Medical Adviser in Plain English or Medicine Simplified*, 11th edition, World Dispensary Medical Association, Buffalo.

Thomson S. 1825, *New Guide to Health; or, Botanic Family Physician*, 2nd ed., For the author, Boston.

Wesley J. 1747, *Primitive Physick: An Easy and Natural Method of Curing Most Diseases* (originally anonymous), Thomas Trye, London.

[7] The manuals included in the corpus are marked with an asterisk.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 67-92
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p67

KNOWLEDGE DISSEMINATION AND EVIDENTIALITY IN THE GENRE OF POSTERS
Anatomy of a condensed medical discourse*

STEFANIA M. MACI
UNIVERSITÀ DEGLI STUDI DI BERGAMO

Abstract – The genre of scientific posters is a very complex one, because it implies combining written and oral modes in communication. Such complexity is further increased by the fact that posters are created in such a way as to stand alone and do the *talking*, while *showing* medical research, all in a single visual plan. Such extreme conciseness is possible only if redundant information, seen as accessory matters, is deleted. As pointed out by Hobbs (2003, p. 459), this means that in the medical context the cohesion usually provided by explicit linkage is supplied by the reader's background knowledge. In this context, the evidential markers, while facilitating the understanding of poster cognitive mapping, indicate the author's level of expertise towards knowledge. Given the fact that consistent linguistic investigations of posters are almost absent from an applied linguistics perspective, it is the aim of this study to describe how evidentiality is realized in such a condensed and specialized genre. More specifically, drawing on Chafe (1986), this investigation will be focused on those linguistic forms regarded as evidential markers and showing various degrees of knowing within the written form of medical posters, in order to illustrate how evidentiality is linguistically realized, and what, if any, pragmatic functions it has. This investigation, based on the analysis of the verbal components of a corpus of 28 medical posters published online between 2002 and 2011, has been carried out on *attested* language use in the written discourse of medical posters. The findings highlight the fact that evidentiality is dependent on the socio-interactional work the speaker does to construct authority, responsibility and entitlement in a particular context and with particular recipients.

Keywords: Knowledge dissemination; Corpus Linguistics; Evidentiality; Medical discourse.

* The research on which this study is based is part of the MIUR-funded research project of national interest (PRIN) *Knowledge Dissemination across media in English: continuity and change in discourse strategies, ideologies, and epistemology* (COFIN grant no. Prot. 2015TJ8ZAS_002), directed by prof. Marina Bondi of the University of Modena and Reggio Emilia – local coordinator, prof. Maurizio Gotti.

1. Introduction

Efficiency in Academia is seen positively as a form of *speed* (Kress 2010), which is ambiguously identified under 'publish or perish'[1] pressure, according to which academic members acquire professional acknowledgment if they publish before others do. In the medical sciences, in particular, such speed identified under 'publish or perish' pressure is indirectly reinforced by the rapid increase of specialised journals, a place where research can be printed and prestige can be acquired, but also by international conferences, a place where research can be 'shown'. For instance, in November 2017, MEDLINE[2] counted 30,000 journals and books for biomedical literature of which 5,278 indexed medical journals;[3] in addition, a Google® search for the phrase 'medical conferences 2017' yielded 530,000,000 'hits'.

As regards conferences in particular, efficiency has often been identified with the organizing of poster sessions, because "[g]iven the limitations of time, the poster format does provide for the maximum number of presentations to be scheduled in a given period, space permitting" (Pearce 1992, p. 1680), which is in line with some data gathered by the author of this paper, namely:

- information obtained from the Research Director of the Publishing Activities Unit of Istituto Superiore di Sanità, the National Health Institute of Italy (http://www.iss.it);
- analysis of the data emerging from the ethnomethodological approach adopted for this investigation (see section 3);
- data collected from an online background survey (http://freeonlinesurveys.com) carried out by the author of this paper (cf. also Maci 2016);
- records obtained from some British and US doctors involved in the academia who were interviewed at the 2011 Meningitis Research Foundation Conference and the 2011 National Cancer Research Institute Conference (see section 3);
- general information about poster sessions provided by previous literature (cf. Dubois 1985a, 1985b; D'Angelo 2010, 2012, 2016a, 2016b; Maci 2011, Maci 2012a; Maci 2012b, Maci 2015a, Maci 2015b, Maci 2016; Rowe, 2012; Rowe 2017).

This is also evident by the fact that most medical conferences have a call for posters only, as, for instance the 10th National Cancer Research Conference (November 2018; cf. https://conference.ncri.org.uk/abstract-submission-2/),

[1] This expression was first mentioned by Wilson (1942).
[2] MEDLINE is the U.S. National Library of Medicine, available at: http://www.nlm.nih.gov/ (08.03.2018).
[3] Data retrieved from PUBMED at http://www.nlm.nih.gov/tsd/serials/lsiou.html (08.03.2018).

where the conference scientific committee selected what poster abstract, amongst all the abstracts received, would be turned into a presentation. Apart from 'logistical' reasons, the reason why medical poster sessions seem to be preferred by researchers lies in the fact that the organisation of poster sessions offers a chance to create relaxed interaction between poster author and audience, from which they can all benefit in terms of scientific support and communication exchange. In addition, the informality created by a poster session forum encourages the development of new research projects. Moreover, although posters must conform to the 'editorial' requirements of the organizing committee, as far as size and format are concerned, authors write creatively in order to attract a potential audience, to inform and persuade. From a more practical point of view, researchers may prefer poster presentations rather than oral ones because the abstract of their poster can be published in major medical journals, which is a convenient means for furthering medical careers. Yet:

> [t]he purpose of medicine is to serve the community by continually improving health, health care, and quality of life for the individual and the population by health promotion, prevention of illness, treatment and care, and the effective use of resources, all within the context of a team approach (Calman 1994, p. 1140).

So, medical research, roughly speaking, is about improving health, the prevention of illness and treatment to let people live well. The only way to find out whether a health treatment works and is safe is to test it. The results of these tests are 'evidence'[4] and medicine based on these tests is called 'evidence-based medicine'. More specifically, evidence-based medicine is

> the conscientious, explicit and judicious use of current best evidence in making decisions about the care of the individual patient. It means integrating individual clinical expertise with the best available external clinical evidence from systematic research. (Sackett et al. 1996, p. 71).

This has nothing to do with the concept of *evidentiality* in linguistics, which is the topic of this paper. Evidentiality in English is a semantic category expressed with lexical markers asserting a factual claim and indicating a source of knowledge (Anderson 1986). The purpose of this paper is that of showing the extent to which, if any, knowledge is linguistically conveyed through evidentiality in the genre of medical posters. The analysis moves from the introduction of the notion of evidentiality and the literature review about evidentiality (Section 2), followed by a description of the methodological approach adopted (Section 3) to a focus on the results both in quantitative and qualitative terms (Section 4), which will be interpreted by looking at the

[4] For a definition of evidence in medicine, see https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4338513/ (08.03.2017).

pragmatic aspects evidentiality has in relation to the poster sections. The concluding remarks (Section 5) will offer a summary and reflection about the role evidentiality has in the genre of medical posters.

2. Evidentiality in linguistics

Boas (1911, p. 43) was possibly the linguist who first, indirectly, introduced the concept of evidentiality by referring to the suffix denoting the source of information in native American Kwakiutl. The term evidentiality, however, was probably coined by Roman Jakobson (1957) as a label for a verbal category indicating the source of information on which the speaker's/writer's statement is based. However, in the mid-1980s a great interest in evidentiality, started with the work of Chafe and Nichols (1986), and continued with Willet (1988), which still seems to be pertinent (DeLancey 1997, 2001; Kamio 1997; Dendale, Tasmowski 2001; Fitneva 2001; Lazard 2001; Nuyts 2001, 2015; Plungian 2001; Aikhenvald, Dixon 2003; Aikhenvald 2004, 2007, 2018; Squartini 2008, 2012, 2018). Nowadays, evidentiality is defined in various ways. The different theories can be grouped into two main approaches, a narrow and a broader definition of evidentiality.

In a narrow definition of evidentiality, "evidentials express the kinds of evidence a person has for making factual claims" (Anderson 1986, p. 273) and include those linguistic markers indicating the source of information, i.e. whether information has been acquired by being seen, heard, inferred or told (Aikhenvald 2003). This approach acknowledges that, in about a quarter of the world's languages, marking information sources is obligatory and that these languages have a grammatical category of evidentiality, while other languages have evidential extensions of non-evidential categories (Aikkhenvald 2007). Yet such a narrow approach to evidentiality considers as evidentiality only certain grammaticalized expressions, in particular morphemes (Aikhenvald 2004, p. 6; Mushin 2001, p. 35). This results in the almost total exclusion of English from such research (Bednarek 2006).[5]

The broader definition of evidentiality is based on the assumption that "[l]anguages typically provide a repertoire of devices for conveying […] various attitudes toward knowledge" (Chafe, Nichols 1986, vii; cf. also Chafe 1986, p. 262), including the source and reliability of people's knowledge. As such, the term evidential "covers much more than the marking of evidence per se" (Chafe, Nichols 1986, vii). In this approach, evidence is just one, but not

[5] This also explains why evidentiality was not studied in languages which do not have grammaticalized expressions of evidential markers until the 1980s. Indeed, the very fact that there are languages that express evidentiality in non-grammaticalized terms made linguists think that, for those languages, evidentiality did not exist.

the only one, of the “epistemological considerations” that are linguistically encoded in the concept of evidentiality (Chafe 1986, p. 262). “Evidentiality in this sense is concerned with matters of truth, certainty, doubt, reliability, authority, confidence, personal experience, validity, inference, reporting, factual and imaginative stance, evidence, confirmation, surprise, and expectedness” (Bednarek 2006, p. 637). The broad categorization of evidentiality offered by Chafe (1986, p. 262) includes the concept of *attitude* toward knowledge, which in Palmer’s wording (1986), is an expression of *modality*. In the relationship between modality and evidentiality, Dendale and Tasmowski (2001, p. 342) identify *disjunction*, *inclusion* and *overlap* approaches. The *disjunction* approach sees evidentiality and modality in opposition, as the former asserts the source of information, while the latter evaluates a proposition (see also De Haan 1999, Aikhenvald 2003, 2004). In the *inclusion* approach, one definition includes the other (cf. Givón 1982, Chafe 1986, Palmer 1986, Willett 1988, Papafragou 2000, Mushin 2001, Ifantidou 2001). In the *overlap* approach, evidentiality and modality ‘intersect’ whenever ‘inductive evidentiality’ is identical to the “modal value of epistemic necessity” (Dendale, Tasmowski 2001, p. 342; cf. also van der Auwera, Plungian 1998, Plungian 2001). According to Dendale and Tasmowski (2001), these three approaches are possible and coexist because there are languages whose evidential systems allow both the sources of information and “the speaker’s attitude towards the reliability of that information” (Dendale, Tasmowski 2001, p. 343; cf. also González 2005, Cornillie 2007).

Any scholar focusing on the concept of evidentiality must decide whether to adopt the narrow or the broader view of evidentiality (Mushin 2001, p. 51). If we take into consideration the narrow definition of evidentiality, we consider just one part of the greater study of epistemological positioning: the aspect of language that marks the speaker’s/writer’s source of information.[6] If we analyse evidentiality in its broader definition, there is much more: we look at the basis of the speaker’s/writer’s knowledge, as well as his/her attitude.

For the purposes of this paper, the type of evidentiality analysed here will follow the broader definition drawing from Chafe (1986) and Mishun (2001), but will not take into consideration the author’s attitude. Indeed, as De Haan (2012) claims, the function of evidentiality is that of asserting the factual claim rather than of evaluating the factual claim (see also Aikhenval 2003).

Chafe’s definition of evidentiality involves several concepts. The first concept relates to the notion of the *mode of knowing*, i.e. the way in which knowledge is acquired. There are four distinctive modes of knowing: *belief*, *induction*, *hearsay* and *deduction*. Each mode of knowing depends on different

[6] I would like to thank the anonymous reviewer who underlined here that the expressions “source of information” and “source of knowing” are used by scholars as synonyms, though normally those (as Chafe does) who use the latter expression also includes the notion of “modes of knowing”.

sources of knowledge: belief and/or opinion are concerned with absence of *evidence* (Chafe 1986, p. 266), induction with presence of *evidence*; hearsay depends on *language*, and deduction on *hypothesis*.

In Chafe's (1986, pp. 266-269) opinion, knowledge deriving from *belief* can be expressed by mental or cognitive verbs such as 'think', 'believe', guess', 'suppose' and so on; when knowledge comes from *induction,* the most common marker is the modal verb 'must', together with such lexical items as 'obvious', 'clear', 'evidently', 'seem'. He further claims (1986, p. 268) that evidence can also denote a *sensory* source of knowledge. In this case, the evidential makers most commonly employed are 'look', 'see', 'hear', 'feel', 'sounds like', 'look like'. When knowledge is acquired through language, its source is based on *hearsay* (Chafe 1986, pp.268-269), normally realized with such verbs as 'say', 'tell', 'suggest', 'apparently', 'seem', 'suppose to'; when metaphorically used with a rumour-like meaning, and in academic discourse, by citation and quotations. *Deduction* is a source of knowledge based on hypothesis (Chafe 1986, pp.269-270), and normally expressed by such markers as 'should' and 'presumably', and, to a lesser degree, 'can', 'could' and 'would'.

Knowledge is information conveyed by the speaker/writer with evidential markers on a continuum of *reliability,* ranging from the most reliable to the least reliable knowledge. The (un)reliable knowledge we acquire from different sources and in different modes is normally checked against what we already know, which in our mind forms categories and frameworks of the world. When there is no match, we have low codability verbally expressed by means of *hedges*; when a match occurs, it is in line with *expectations*. According to Chafe (1986, pp. 264-265) *reliability* is expressed by means of epistemic expressions (adverbs, adjectives, modals etc.) indicating certainty, closeness or remote possibility; as well as epistemic modals. *Hedges* are present whenever we try to match the categories and frames of the world we already know with what we are learning. This tentative match is normally expressed with nouns, verbs, adjectives or predications such as 'sort of', 'about'. If, on the other hand, a match is possible, we have *expectations,* lexically realised with such expressions as 'of course', 'in fact', 'actually', in general, with connectives or adversatives, "which relate in various way to expectations" (Chafe 1986, p. 271). Yet, as Mushin (2013, p. 634) claims, although there is no link between evidentiality and evaluation, between source of knowledge and reliability, or attitude, in Chafe's term, it is true, however, that the selection of evidential markers by the speaker/writer may indicate a form of evaluation. "In other words, grammatical evidentials are a resource that speakers use to express their stance towards their knowledge" (Mushin 2013, p. 635). In De Haan's (1999, 2001, 2005, 2012) opinion, however, evidentiality cannot be used to reveal attitude and evaluations, because its main function is that of validating a factual claim (Anderson 1986), rather than the speaker' attitude.

A summary of Chafe's evidential model can be seen in Table 1, below:

EVIDENTIALITY		
(un)reliable ***knowledge*** (un)matching with speaker's/writer's previous experience (hedges/boosters/)expectations		
language	$\longleftrightarrow$	Hearsay
lack of evidence	$\longleftrightarrow$	Belief
sensory/cognitive evidence	$\longleftrightarrow$	Induction
hypothesis	$\longleftrightarrow$	Deduction

Table 1
Adapted from Chafe's (1986, p. 263) notion of evidentiality.

Clearly, in discourse, knowledge (un)reliability and matches against previously acquired experience overlap with knowledge source and mode; in contrast, knowledge source and knowledge mode are inextricably dependent one upon the other.

2.1. Poster Literature review

Posters can be realized in varying sizes, contain varying numbers of tables and graphs, may have neither references nor abstracts. Nevertheless, medical posters in their written form have an extreme standardization of their narrative pattern.

The narrative pattern they describe is always vertical and the reading path offered is strategically indicated according to the various sections, called, as we will see later, *Introduction*, *Methods*, *Results* and *Discussion* (cf. for instance, Swales 1990; Gross et al. 2002, and also Cargill, O'Connor 2010) whose headings are always signposted in the text. Therefore, by following the labels, the reader can construct the correct reading pattern.

Although there is a large amount of literature and guidelines available on scientific posters by medical authors (cf., for instance, AIFA 2005; American Heart Association 2018), from the perspective of applied linguistics, the genre of posters has undergone little investigation. The first description of posters was offered by Dubois (1985a, 1985b), who examined the generic features of posters and the ways in which they are presented. She underlined their main function of *popularizing* scientific communication by exploiting elements used to attract a (medical) professional, as well as a layman, audience in order to create potential networks amongst research teams. Swales and Feak (2000, p. 81) suggest that for many years posters have been "the poor country cousin of papers, but recently they have gained in status"; their approach is mainly pedagogical and aimed at helping novice writers to produce effective posters, defined by the authors as the public display of academic writing (conclusions which are also reached by Burgess, Fagan 2004). Swales (2004, p. 21) considers the genre of posters to be a hybrid form falling between "elements of the research paper and conference visuals or handouts". Indeed, he sees posters as a multimodal communicative event, with text, graphics, colour and (interactive) speech used to convey meaning. On the other hand, MacIntosh-Murray (2007) claims that learning to communicate scientific knowledge through posters is about much more than mastering poster fonts, colours and sizes: posters form a complex genre because of the multiple role they play in both written and oral communication, which has to be consistent with (a) professional, editorial and generic constraints, (b) interaction with an expected audience and (c) professional prestige. In addition, posters are meant to stand alone, without the presenter. In other words, posters are supposed to *do* the talking (MacIntosh-Murray 2007, pp. 351-352) and *show* medical research. As far as we know, the only contributions about posters are those made by D'Angelo (2010, 2012, 2016a, 2016b) and Maci (2011, 2012a, 2012b, 2015a, 2015b, 2016). While D'Angelo is interested in the ways in which multimodal interactional and interactive strategies are exploited to guide the readership through verbal and visual elements characterizing posters in a dialogic interaction with the poster's author, Maci focusses her attention on the poster genre, its diachronic development as well as on syntactical and multimodal aspects of poster generic structure.

To the best of our knowledge, no study has ever considered the way in which evidentiality intersects with medical discourse in posters. In other words, the aim of this paper is to investigate the extent to which knowledge acquired by a medical researcher is linguistically realized in terms of source of information in medical posters. More precisely, the investigation will focus on the following research questions (RQs):

- What are the evidential markers whose mode is defined in terms of hearsay, belief, induction and deduction and used to show various degrees of knowing?
- What evaluative/pragmatic functions do they have, if any?

The analysis of evidentiality will be drawn on Chafe (1986) and De Haan (1999, 2001, 2005, 2012). In order to describe how evidentiality is realized in the genre of posters, it seems necessary to investigate *attested* language use rather than assuming any theoretical implications applicable to them. For this reason, we will also draw on corpus linguistics. Broadly speaking, corpus linguistics is "the study of language based on examples of real life language use" (McEnery, Wilson 1996, p. 1). It is an empirical approach, since the enquiry starts from authentic data and aims to analyse and describe language use as realised in texts (Tognini-Bonelli 2001, p. 2). Consequently, the study carried out to answer the RQs may qualify as a corpus-based approach, in that it relies on both quantitative and qualitative techniques (Biber, Conrad 2001; Mair 1991).

3. Methodological approach

Following an on-line background survey carried out with the help of the on-line medical journal *il Pensiero scientifico* and some interviews made at medical conferences (namely the 2011 Meningitis Research Foundation Conference and the 2011 National Cancer Research Institute Conference), the following sites were suggested, from which posters were randomly selected, as indicated in Table (2) below:

Websites	**No. of posters**
Istituto Superiore di Sanità (http://www.iss.it)	3
Barts and The London NHS Trust (https://web.archive.org/web/20160423000317/http://www.ihse.qmul.ac.uk/cme/bscmeded/poster/index.html	1
New York City Health and the Mental Hygiene Department (https://web.archive.org/web/20120114004140/http://www.nyc.gov/html/doh/html/dires/epi_posters.shtml)	7
International Conference on Retroviruses and Opportunistic Infections (CROI) (http://www.croiconference.org/abstracts/search-abstracts/)	7
2011 International Conference on Meningitis (http://www.meningitis.org/posters)	1
Eposternet (http://www.eposters.net/)	3
F1000 (https://f1000research.com/f1000posters-message)	2
Posters uploaded (www.slideshare.net)	4
TOTAL	**28**

Table 2
Breakdown of poster selection.

Twenty-eight posters were downloaded from the above-mentioned websites (41,587 words). Copyright permissions to use the posters collected and published online have been granted by all the institutions and journals listed above. All downloaded posters, in secure pdf. and in jpg. formats, were converted into word format with Abbey Transformer, an OCR software, and then saved as .txt to allow analysis with Wordsmith Tools 6 (Scott 2012).[7]

An already established theoretical framework was used, following the broader definition of evidentiality by Chafe (1986), and drawing from Chafe's categories to classify data. Yet the approach used was also *text-driven*, similar to Tognini-Bonelli's (2001) concept of corpus-driven linguistics.[8] Thus, the

[7] "WordSmith Tools is an integrated suite of programs for looking at how words behave in texts." Available at https://lexically.net/downloads/version7/HTML/index.html # (15.11.2018).

[8] According to Tognini-Bonelli (2001), corpus-based linguistics is a linguistics that uses a corpus to explain a theory, whereas corpus-driven linguistics is a linguistics that formulates a theory on the basis of a corpus.

exploitation of any preconceived classification of evidential markers was avoided (as, for instance, Aikhenvald 2004), which might bias the investigation. Indeed, already-existing categorizations of evidentials are certainly valid for the corpora they are extrapolated from, but they may give completely different results when applied to other specialised corpora such as the medical corpus; for instance, certain evidentials which can occur in the genre of medical posters may be disregarded in other genres simply because the evidentials here detected may not occur in those genres or in other scholars' theoretical classifications of evidentials or because they may be realized with lexical markers that are not present in those genres (and vice versa). For this reason, a sample sub-corpus from the medical poster representative corpus needed to be created in order (a) to have random samples of the types of evidential markers most commonly occurring in medical posters, and (b) to support claims regarding the frequency and distribution of these features (i.e. types of evidentials). Indeed, the study was preceded by the manual analysis of a small-scale text sub-corpus, rather than by an approach based on automated large-scale corpus analysis. For this reason, five posters were randomly selected out of the above described and indicated in Table (2), above, and read in order to detect potential evidential markers, without categorizing them on an a priori basis. The markers were first identified according to the *source* of knowing. In other words, the first step was to detect whether there was a source of knowing or not, or whether the author of the text was using language as the source of knowing or forming a hypothesis. For instance, in (1) below, the item *thought* represents an evidential marker with no source of knowing (my emphasis):

> (1)This inhibition is <u>thought</u> to be mediated through the cholesterol-rich domains on the cell surface (lipid rafts) through which HIV-1 emerges during viral maturation. […] Since all cholesterol is provided by the host, and since cholesterol is important to the viral life cycle, it is reasonable to hypothesize that HIV may have evolved to produce viral-host gene interactions that will up-regulate intracellular cholesterol. (P080)

Indeed, the passive form of the evidential form *is thought* introduces a factual claim that, in Chafe's terms, expresses a lack of evidence which has to be tested (as confirmed by the second part of the excerpt starting with "it is reasonable to hypothesize that…"). Not only is evidence missing, but also the source of information, as the absence of the agent reveals. Although in medical discourse, and in the hard sciences in general, this is a strategy employed to allow the focus on processes rather than on people (cf. Halliday, Martin, 1993), the absence of the agent is a way to avoid a face threat (Hyland, 2017). Indeed, the subject of the factual claim "inhibition" is the patient of the main clause which has occupied the subject position in a process called object-to-subject raising (Quirk et al. 1985, p. 731), thus creating vagueness in the source of knowledge attribution.

In (2) below, the source of knowing is specified by the verb *demonstrate* which refers to the evidence offered by "data":

> (2)These data demonstrate the presence of substantial levels of non- and mono-sialylated core 1 O-linked carbohydrate on the gp120s of SIVmac and SIVsm [...] (P483)

In other cases, the source of knowing is offered by language, as in excerpt (3):

> (3)[…] all reported partners are true sexual, needle sharing, or social network partners (P272).

In others, the source of knowing could be a hypothesis, as in (4):

> (4)This suggests the anti-bacterial properties of RMGIC and GIC should be equally effective. (P054)

As can be seen, excerpt (1) can be categorized as 'lack of evidence', while excerpts (2)-(4) as 'evidence' realized via 'data', 'language', or 'hypothesis', respectively. Following Chafe (1986), the lexical evidential markers thus found were then grouped according to the mode of knowing as well as according to the *belief*, *induction*, *hearsay* and *deduction* categories that (1), (2), (3) and (4), respectively, belong to.

On the basis of the evidentials found in the five randomly selected corpus, the analysis of 28 posters allowed the detection of 2,158 potential evidential markers. Each evidential marker was run in Wordsmith Tools 6 to check concordancing, and each concordance was then manually checked to see whether the item was a real evidential or not. Verbs were also checked in their conjugated form, even if an item was not present in the corpus. For instance, if the item *estimate* was present in the corpus, the conjugated forms *estimates* and *estimated* were 'searched' for as well. Once all the concordancing had been checked, the items were classified according to Chafe's (1986) interpretation and grouped into *hearsay, belief*, *induction* and *deduction* categories. These were then compared with semantic domains computed by WMatrix, which permitted[9] the extraction of key domains by applying keyness calculations (a log-likelihood test) to tag frequency lists. The combination of keywords and key domains offered by WMatrix and the evidentials allowed the detection of further evidential markers which may have been missed. They, too, were double-checked, first with WordSmith Tools for concordancing lists, and then manually.

[9] WMatrix, a software offered by Lancaster University, is a type of automatic tagging software that assigns part-of-speech and semantic field (domain) tags.

A closer analysis of our data allowed the identification of three subcategories that were not relevant for this study. First, a large group of evidentials was actually formed by domain-specific words.[10] For instance, the term *significant* in excerpt (5) indicates a 'statistical significant correlation between two elements' rather something that is relevant, while the term *interaction* in (6) means 'mutual or reciprocal action or influence' from a genetic angle:

(5)[…] Compared to the MIC distribution of the previous year, no <u>significant</u> shift of the MIC values of vancomycin was observed. (P052)

(6)The HIV-host interactions were visualised as a protein <u>interaction</u> network. (P283)

Second, some words were evidential markers but metadiscursively so, in the sense that they guided the reader within the text in order to find information referred to. There are such poster headings, such as *introduction*, *conclusions* and *methods*, and labels, such as *figure*, *table* etc., of which we can find some examples of their contextual use below:

(7)<u>Introduction</u>
MolPAGE (Molecular Phenotyping to Accelerate Genomic Epidemiology) is [...] (P155)

(8)<u>METHODS</u>
This is an epidemiological study in which we assessed the clinical characteristics of all patients with […] (P333)

(9)<u>Conclusions</u>
*Density mapping enabled identification of a previously unrecognized geographic focus of HIV […] (P478)

(10) <u>Figure</u> 1: Data Flow in […] (P272)

(11) <u>Table</u> 1. Summary of Simian Retrovirus Serology. (P002)

Last, there are items that have been defined as 'co-textual' evidentials. They are items that better specify the type of evidence that the authors have at their disposal. An example can be seen in example (12) below:

(12) HCV <u>laboratory</u> results. (P001)

[10] All the words were semantically checked with the *Merriam-Webster* medical dictionary at http://www.merriam-webster.com/medical/interaction.

As both metadiscursive and co-textual evidentials reveal, the source of information does not validate the factual claim, thus they have not been regarded as strictly evidentials for the purposes of this paper. Indeed, they do not express the degree of knowledge reliability indicated by the author of the text, according to Chafe's (1986) framework.

Of the remaining evidentials, only those which occurred more frequently than five times were taken into consideration. This is because we observed that an item occurring five times or more is less likely to appear on one poster only, which may also convey less skewed results. We were thus left with 125 evidential occurrences.

This quantitative analysis was then followed by a qualitative one (Coffey, Atkinson 1996; Miles, Huberman 1994), which allowed interpretation of the findings of this study.

4. Preliminary results

4.1. Evidential Items and their frequency

The application of the methodological approach described in section 3 allowed the identification of the following evidentials:

- belief: think;
- deduction: estimate;
- hearsay: according to, confirm, explain, predict, describe, report, publish, recommend, suggest, scholarly quotation and endnotes;
- induction: demonstrate, indicate, result, show, detect, determine, indicate, investigate, observe, see, show, represent.

As explained in section 3, above, only those evidentials occurring more than 5 times have been taken into consideration for the investigation. Since *think* and *estimate* are hapax, they are dismissed; the investigation has, thus, focussed only on the category of hearsay and induction evidentiality. Table (3) shows the raw frequencies of evidentials grouped according to the category they belong to (indicated in rows two to four) and a number introduced in the table by the heading STTR, which indicates the Standardised Type/Token Ratio, i.e. occurrences normalised to a text length of 1,000 words (Hunston 2002):

Evidential	Raw frequency	STTR
Belief	//	//
Deduction	//	//
Hearsay	56	13.96
Induction	69	15.59
TOTAL	**125**	**23.82**

Table 3
Frequency of evidentials.

The analysis of evidentials in the corpus of 28 posters shows variation in their frequencies. Given that the evidentials taken into consideration are those which occur at least five times (see section 3, above), the table per se suggests that the most remarkable features are: (a) the absence of belief evidentials and deduction evidentials; and (b) the high frequency of inductive and hearsay evidentials. As to (a), the absence of belief evidentials may probably be justified by the fact that elaboration of the author's argumentation cannot be done in the genre of posters, as it belongs to a genre which has space constraints and is therefore devoted to in-progress research (cf. Maci 2011, 2012a, 2012b, 2015a, 2015b, 2016); it is often only later, in a paper, that the author can argumentatively elaborate a theory (as to the use of belief verbs, see, for instance, Thomas, Hawes 1994; as to how argumentation is developed in research articles, see, for instance, Nwogu 1997, Maci, 2012a, 2012b).

Also, the absence of deductive evidentials is predictable – medicine is science-based with an inductive bottom-up approach: experiments and observation may lead to a tentative hypothesis which can be further tested to yield a final theory.

Similarly, the presence rate of hearsay evidentials does not come as a surprise. Indeed, as confirmed by Biber et al. (1999, p. 372):

> [a]cademic prose reports relatively few physical, mental, or communication activities – and when such activities are reported, they are often attributed to some inanimate entity as subject of the verb.

As we can see in excerpt (13), the hearsay evidential has *studies* as a source of information:

> (13) The genetics studies and immunology studies underlined genetic predisposition combined with environmental factors playing the major roles in the etiology of autism. (P274)

Clearly, since studies cannot *suggest*, the hearsay verb has to be meant metaphorically, thus overlapping with inductive evidentials.

An examination of the semantic domain grouping hearsay evidentials, elaborated with WMatrix, which can be seen in Table (4) below, led us to consider that scrutiny of the distribution of *all* evidentials seems necessary across the *Introduction*, *Methods*, *Results*, and *Discussion* (IMRD). IMRD sections of the posters under investigation:

1	of both diseases - The CDC	recommends	that all IDUs be tested for HC
2	; Objectives - To	describe	rates of HCV testing and infec
3	at the high end of the range	reported	for MSM in the United States (
4	Background : Current evidence	suggests	that HIV-1 and HIV-2 have orig
5	Current evidence	suggests	that HIV-1 and HIV-2 have orig
6	; cell associated . Here , we	report	data from a CDC linked study i
7	by ELISA and WB as previously	described	by Lerche et al . SFV antibodi
8	imers were used as previously	described	(Heneine et al, Lerche et al
9	NHPs (14 yrs with chimps) ;	reported	bites , scratches but does not
10	were consistently negative ,	suggesting	low-level viremia . Persistent
11	ity was observed for Case 1 ,	suggesting	a persistent infection . Case
12	erforms invasive procedures ;	reports	bite-scratch injuries to macaq
13	bility All SFV-infected cases	report	being in good health with the
14	V , or STLV and may likely be	explained	by the higher frequency of exp
15	mented exposure of 14.5 years	suggest	that secondary transmission by
16	rkers will be needed to fully	define	the pathogenic potential and h
17	two SRV-D seropositive cases	suggests	that cross-species infection w
18	/oxytoca : 743 isolates-15.2%	reported	as ESBL positive , 8% multires
19	w , but high compared to what	reported	in other European countries .

Table 4
Hearsay Semantic domains.

4.2. Evidentials distribution across IMRD poster sections

Previous research has revealed that medical posters have a highly codified pattern which, while transcending national constraints (Dahl, 2004, p. 1822), follows the constraints described in the *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*,[11] originally known as Vancouver style, drafted in 1978 by the *International Committee of Medical Journal Editors* (ICMJE) and published in 1991 (last update, December 2017) in the *British Journal of Medicine* (BJM). Such a codified pattern adopted by poster authors is identical to that of other scientific writing, and has an *Introduction* (and *Objectives*), *Methods*, *Results*, and *Discussion*/*Conclusions*.

A close examination of evidentials across the poster IMRD pattern revealed an interesting outline, of which Chart (1) gives the major features. We

[11] Further information at http://www.icmje.org/recommendations/ (01.10.2018).

can see here that evidentials are predominantly used in the *Results* section, and yet, if we look at single evidential categories, their distribution varies according to the section they occur in. For instance, induction evidentials are more frequent in the *Results* and *Discussion* section than in the *Introduction* section, while hearsay evidentials have roughly the same frequency in the *Introduction* and *Results* sections, which is much higher than the *Methods* and *Discussion* ones.

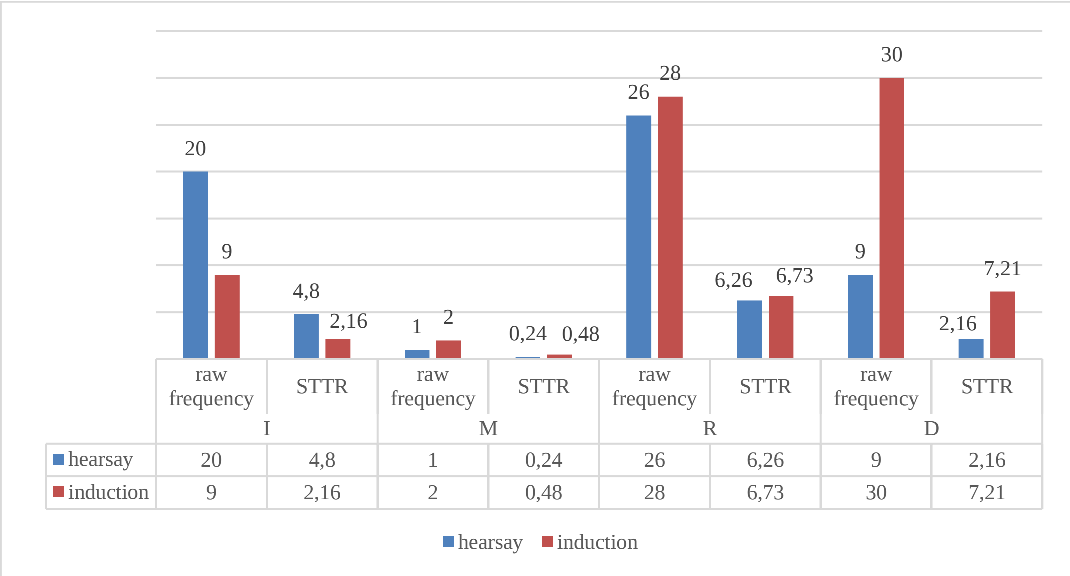


Chart 1
Distribution of evidentials across poster IMRD pattern.

Furthermore, within each category, we notice that evidentials are not randomly used but rather they follow a precise pattern. In other words, there is a preference for using certain evidentials in particular sections. For instance, the evidential *show*, which is an evidential marking an inductive (through perception) source of information, is more likely to be found in the *Discussion* section than in other sections.

In the following paragraphs, we will see in detail how the various evidentials markers are distributed in the IMRD sections of posters.

4.2.1. The Introduction *section*

In poster genre, the *Introduction* section is normally very short. It introduces background information, to which a problem or an issue is related and, finally, the explicit aim of the poster itself. All this is realized in very concise and elliptical language. Here, evidentials (30 occurrences, STTR 7.21) emphasize these points featuring the *Introduction*.

Hearsay evidentials (20 hits, STT 4.80), are normally realized in three different ways. They can focus on published literature reporting previous research or current evidence (my emphasis, in bold) and which can be expressed by an adverbial expression, as in (14):

> (14) According to previous studies, antimicrobial resistance of S. aureus and S. pneumoniae seems to be stable. (P052)

Hearsay evidential markers can make reference to quotation to scholarly literature, as in (15), or footnotes, as in (16). Indeed, differently from Aikhenvald (2004, p. 64), Chafe (1986, pp. 268-269) underlines that "in academic writing knowledge obtained through language is indicated with the formal devices of citing a reference or personal communication", which is exactly what happens in medical posters.

> (15) Training activities on environmental and occupational health issues and work safety (**Quito, Guayaquil 2008**). (P337)

They can also be realized as verbal expressions, as the expression "indicating that" in (16), below:

> (16) [GSTPl was consistently down regulated or not expressed in prostate cancer, which coincides with previous research **indicating that** [GSTPl is methylated in prostate cancer tissue **[16,95,99-109,111-118,130,136]** (P476)

Induction evidentials (9 hits, STTR 2.16) found in the *Introduction* section are defined as knowledge deriving from evidence and these coincide with lexical items which are mainly referred to as *data*, *diagnoses*, *evidence*, *investigation*, *research* and *results*, that is, with all those elements which in medicine are at the basis of inductive scientific methods: observation, experiments, data test and analysis, as excerpts (17) and (18) below seem to indicate:

> (17) We previously **demonstrated** that some, though not all GBV-C NS5A proteins inhibit PKR-mediated eIF-2a phosphorylation, and this may help the virus avoid clearance by cellular antiviral response mechanisms (4). (P165)

> (18) Observational, clinical, and laboratory evidence **indicate** [sic] that sex steroid hormones are important to the development and progression of prostate cancer [2-14] (P476)

4.2.2. The Methods *section*

The *Methods* section of posters explains the features characterizing the scientific experimental protocol (subjects, procedures, statistics analysis, ethical approbation). It is a compulsory section and the most important one, because it is on the basis of the methodological approach of the research or experiment that the whole study is scientifically evaluated. In other words, the description of the methodology does not need any source of information in evidential terms.[12] This explains why only three evidentials (STTR 0.48) are used in the *Methods* section: one hearsay evidential, and two inductive evidentials.

The hearsay evidential clearly indicates in detail the procedure/experiment followed/conducted, questionnaires and records, and reports sources of information:

> (19) Cell surface receptor density: CXCR4 and CD4 expression on Jurkat cell surfaces were characterized by flow cytometry as previously **described** (3) (P156)

The two induction evidentials underline how things will be identified; who amongst the diagnosed people have been selected for the experimental protocols; and with what tests, as we can see from excerpt (20) below:

> (20) The Durham Fidelity Criteria specifies [sic] that EI teams should: […] 7. Monitor DUP and collect data to **demonstrate** its effectiveness in relation to key outcomes including engagement rates, relapse rates, hospital readmission, suicide and Para suicide, education and employment functioning. (P398)

4.2.3. The Results *section*

In the poster genre, the *Results* section is the place where data collected from the scientific protocol are generally presented in figures, tables and graphs. In this section, the textual element of posters is strictly integrated with the visual one, the latter normally occupying 30-50% of the available space, as shown by Maci (2016), which means that the verbal and visual components of posters are strictly integrated. Here the use of visuals seems to be preferred because images help readers better understand key points (Mitrany 2005, p. 115). The text used in this section has, therefore, an explanatory function and is mainly realized as captions and, at the same time, conveys key scientific information. The main aim of the *Results* section is objectivity and effectively *seeing* the results expressed in graphs, tables and figures persuasively, which assures readers that the scientific procedure has been correctly followed and that the protocol can be tried and tested in any laboratory under the same conditions, as described on the poster. In addition, the presence of tables and figures demonstrates that the evidential truth represented here is unambiguous (Skelton, Edwards 2000, p. 1268). This explains why belief and deduction evidentials are absent. There is, on the contrary, a high frequency of hearsay evidentials (26 hits, STTR

[12] There might be sources in the methodological section, but this does not mean that they are evidentials. An evidential validates a factual claim by possibly indicating a source of knowledge. For instance, the sentence “The questionnaire was based on the NIMHE National EI audit of February 2005 (Pinfold, V., Smith, J. & Shiers, D., 2007).” has a source of information but no evidentiality, as there is no factual claim to validate.

6.65), when compared to the other sections. Interestingly, the most frequently used hearsay evidentials found here are verbs that report findings in a process of knowledge construction with reference to previous research as we can see in (21), below:

> (21) RESULTS […] We **have recently reported** that it is possible to label individual particles of HIV using a Vpr:IN-GFP fusion protein (Alberto Albanese, Daniele Arosio, Mariaelena Terreni, and Anna Cereseto, HIV-1 Pre-Integration Complexes Selectively Target Decondensed Chromatin in the Nuclear Periphery. PLoS ONE. 2008; 3(6): e2413. Published online 2008 June 11.) (P340)

The excerpt in (21) above clearly indicates a hearsay evidential as the verb *report* has as a source of evidence *we* and validates the factual claims introduced by *that*. Furthermore, the evidential *reported* is supported by the reference in literature within brackets, which Chafe (1986, pp. 268-269) classifies as hearsay.

There are also, but less frequently, verbs in which data are given the author's voice and explain the meaning of the finding, as in (22):

> (22) RESULTS. […] Figure 2. The initial linear relationship with square root of time **suggests** a diffusion controlled process. (P054)

A similar occurrence of inductive evidentials (28 hits, STTR 6.73) is found here: in our corpus, this type of evidentials is normally verbs whose subjects are not the authors of the poster but rather data and findings. In the poster *Results* section, in particular, the poster authors never seem to have any active involvement in data analysis (see also Gross et al. 2002; Vihla 1999). Graphs, figures and tables *show* or *represent* results. It is the poster that presents the medical research, rather than the poster author. The author's voice does not need to support facts which can speak for themselves and visual components that can stand alone. This is clearly an attempt to avoid both negative face threats (in order to receive consensus from the scientific community) and future research attacks proving the present research wrong:

> (23) Figure 2. This graph **shows** fluoride release from the GIC and RMGIC are comparable. The initial linear relationship with square root of time **suggests** a diffusion controlled process. (P054)

> (24) Panel C and D **illustrate** that NS5A has two immunoreactive bands representing different phosphorylation forms, and that NS5A expression is shut down when cells are grown in a tetracycline (doxy) (P156)

4.2.4. The Discussion *section*

The *Discussion* section of posters outlines a summary of the main results and, at the same time, offers an interpretation of the main findings. Of course, data interpretation is based on the findings resulting from the experiments explained in the *Methods* and stemming from the *Results* section – which, as our data suggest, is based on no belief evidentials but rather on hearsay and induction evidentiality (cf. Table 3, above).

In this section of the posters, the type of language is seldom argumentative and is mainly realized through bulleted sentences, due to space constraints. Conclusions are normally expressed in the present tense, describing a reality which is less likely to be counter-claimed since it is offered as a form of truth belonging to commonly shared knowledge. There is a going back and forth between what has been described in the *Results* section and the interpretation leading to possible theorization of a new scientific protocol. In this context, hearsay evidentials (9 hits, STTR 2.16) are used by the researchers to show their active involvement as a team:

> (25) Resistance to amicoglicosides and cephalosporines in E. coli is still low, but high compared to what **reported** in other European countries. (P052)

> (26) Comparison with independent sources, such as the Medical Monitoring Project, **confirms** that HEFSP captures 90-99% of confirmed HIV+ individuals. (P400)

On the other hand, induction evidentials (30 hits, STTR 7.21) are used as if there is no active involvement of the poster researchers, exactly as in the *Results* section: what has been seen, observed, identified and tested in the *Results* section, may or may not confirm what has been previously reported, following an inductive reasoning path. What is underlined is that the findings, which have been seen, observed, identified and tested in the *Results* section, offer an answer to the RQs previously indicated:

> (27) CONCLUSION […] This **reveals** that the nuclear transport of HIV in the nucleus is an active process similar to transport in the cytoplasm. (P340)

> (28) CONCLUSIONS […] From this study, FTIR **showed** practically 100% polymerisation of this monomer. The fluoride release level from the RMGIC was found to be comparable to that of GIC. This **suggests** the anti-bacterial properties of RMGIC and GIC should be equally effective. (P054)

5. Conclusion

This paper focused on those linguistic forms regarded as evidential markers in order to determine if, and to what extent, knowledge is expressed in terms of belief, induction, hearsay and deduction, following Chafe's (1986) framework, which replies to the first research question posed in this paper, that is:

- What are the evidential markers whose mode is defined in terms of hearsay, belief, induction and deduction and used to show various degrees of knowing?

More specifically, the aim was that of detecting whether the use of evidentiality has a pragmatic function in the genre of medical posters – which provided an answer to the second research question here presented, that is:

- What evaluative/pragmatic functions do they have, if any?

From the analysis, the following points emerged:

i) what evidentials are used in the genre of medical posters;
ii) the distribution of evidentials in poster genre indicates that there is variation;
iii) induction evidentials are used more frequently than other evidentials;
iv) induction evidentials are used more frequently in the Results and Discussion sections of medical posters;
v) hearsay evidentials are more frequent in the Introduction section;
vi) belief and deduction evidentials are rarely used in the poster genre.

Apparently, the distribution of evidential markers responds to the pragmatic function of posters.[13]

In the *Introduction*, posters commonly start with the indication of a research niche, which is only possible if literature on the topic is displayed (hearsay evidentials).

The *Methods* section describes protocols, processes, subjects, ethical probation, in detail, as reported by the poster author who by using hearsay evidentiality, makes references to the medical literature necessary to demonstrate the experiment (induction evidentials).

[13] Although posters have IMRD sections exactly as the genres of RAs and abstract (cf. Swales 2004; Nwogu 1997), and even though bot RAs and abstracts are organized in such a way as to realize IMRD pragmatic functions, it is not possible: 1) to compare posters and abstracts given that no analysis of evidentiality has ever been carried out on evidentiality in the genre of abstract; 2) to compare posters and RAs because the only study carried out on medical RAs and evidentiality (Mocini 2015) focusses on the relation existing among attribution, the source of information, the types of information this source provides, and how the evidence is validated by adopting a Systemic Functional Linguistics approach and the Appraisal Theory and without taking into consideration the IMRD structure of RAs.

The *Results* section, featuring lots of visuals, which in most cases do not need any support from text, exploits evidentials to illustrate (by means of induction evidentials) results in a process of knowledge construction and to link them discursively to the background medical literature (expressed by using hearsay evidentials) cited by the poster author.

The *Conclusion* section presents the people metonymically (hearsay evidentials) responsible for the main results of the investigation (induction evidentials): this is expressed in evidential terms via hearsay and induction evidentiality, the former used to introduce scholarly quotation, the latter to bestow results.

As we are aware of the fact that this investigation is based on 28 posters only, and that posters are extremely short in their content, we will carry out future research with a larger corpus, to test what we have found in this study. This will help us to see whether there are certain lexical and syntactical regular patterns of evidentiality in specialised contextual use (collocation and clusters); furthermore, a follow-up on such a topic will offer better insights into how these different evidential modes rhetorically merge in scientific discourse.

Bionote: Stefania M. Maci is Associate Professor of English Language and pro Vice-Chancellor (Education) at the University of Bergamo. She is the national coordinator of the TRAILs Project funded by the EU. She is member of CERLIS (Research Centre on Specialized Languages), CLAVIER (The Corpus and Language Variation in English Research Group), BAAL (British Association of Applied Linguistics), AAAL (American Association for Applied Linguistics), and AIA (Associazione Italiana di Anglistica). Her research is focussed on the study of the English language in academic and professional contexts, with particular regard to the analysis of tourism and medical discourses. Her most recent publications include: the monographs *The MS Digby 133 Mary Magdalene. Beyond scribal practices: language, discourse, values and attitudes* (2018); *Tourism Discourse: Professional, Promotional, Digital Voices* (2013); the special issue of *Scripta Manent*, edited with Michele Sala and Šarolta Godnič Vičič, *The language of tourism* (2018); the volumes, edited with Maurizio Gotti and Michele Sala, *Ways of Seeing, Ways of Being. Representing the Voices of Tourism* (2017) and *Insights into Medical Communication* (2015); and the papers: '#icantbreathe. Ideology and Consensus: Printed News vs. Twitter. *Critical Approaches to Discourse Analysis Across Disciplines*' (2018); 'An Introduction to English Tourism Discourse' (2018); 'Meaning-making in Web 2.0 Tourism Discourse' (2017); 'Perception or Perspective? Adjusting the representation of Italy and the UK for the tourist: the *Made in Italy* and *This is Great Britain* campaigns' (2016); and 'A framework for medical poster multimodal analysis: a pilot study' (2016)

Author's address: stefania.maci@unibg.it

Acknowledgement: We would like to thank the anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of this paper.

References

AIFA. Ministero della Salute. 2005, *Il poster congressuale: come prepararlo, illustrarlo e renderlo efficace*, in "Bollettino d'informazione sui farmaci" 5-6, pp. 237-240.

Aikhenvald A. 2003, *Evidentiality in typological perspective*, in Aikhenvald A. and Dixon R.M.W. (eds), *Studies in Evidentiality*, John Benjamins, Amsterdam/Philadelphia, pp. 1-32.

Aikhenvald A. 2004, *Evidentiality*, Oxford University Press, Oxford.

Aikhenvald A. 2007, *Information source and evidentiality: what can we conclude?*, in "Rivista di Linguistica" 19 [1], pp. 209-227.

Aikhenvald A.Y. 2018, *The Oxford Handbook of Evidentiality*, Oxford University Press, Oxford.

Aikhenvald A. and Dixon R.M.W. (eds), 2003 *Studies in Evidentiality*, Benjamins, Amsterdam/Philadelphia.

American Heart Association 2018, *Poster Guidelines*. https://professional.heart.org/idc/groups/ahamah-public/@wcm/@sop/@scon/documents/downloadable/ucm_481677.pdf (09.03.2018).

Anderson L.B. 1986, *Evidentials, Paths of Change, and Mental Maps: Typologically Regular Asymmetries*, in Chafe W. and Nichols J. (eds), *Evidentiality. The Linguistic Coding of Epistemology*, Ablex Publishing Corporation, Norwood, New Jersey, pp. 273-312.

Bednarek M. 2006, *Epistemological positioning and evidentiality in English news discourse: A text-driven approach*, in "Text & Talk" 26 [6], pp. 635-660.

Boas F. 1911, *Introduction to the Handbook of the American Indian Languages*, Government Printing Office. https://archive.org/details/introductiontoh00boasgoog (09.03.2018).

Biber D., Johansson S., Leech G., Conrad S. and Finegan E. 1999, *Longman Grammar of Spoken and Written English*, Pearson Education, Harlow, England.

Biber D. and Conrad S. 2001, *Quantitative corpus-based research: Much more than bean counting*, in "TESOL Quarterly" 35 [2], pp. 331-336.

Burgess S. and Fagan A. 2004, *Genre Analysis and the Novice Researcher: The Book Review and the Poster Presentation*, in Brito M. and Oliva J.I. (eds), *Traditions and Innovations: Commemorating Forty Years of English Studies at ULL (1963-2003),* RCEI, La Laguna, pp. 61-77.

Calman K. 1994, *The profession of medicine*, in "British Medical Journal" 309, pp. 1140-1143.

Cargill M. and O'Connor P. 2010, *Writing scientific research articles: strategies and steps*, in "Aquaculture International" 18 [4], pp. 705-706.

Chafe W. 1986, *Evidentiality in English Conversation and Academic Writing*, in Chafe W. and Nichols J. (eds), *Evidentiality. The Linguistic Coding of Epistemology*, Ablex Publishing Corporation, Norwood, New Jersey, pp. 261-272.

Chafe W. and Nichols J. (eds) 1986, *Evidentiality. The Linguistic Coding of Epistemology*, Advances in Discourse Processes, Ablex Publishing Corporation, Norwood, New Jersey.

Coffey A. and Atkinson P. 1996, *Making sense of qualitative data: Complimentary research strategies*, Sage, Thousand Oaks, CA.

Cornillie B. 2007, *Evidentiality and epistemic modality in Spanish (semi)auxiliaries. A cognitive-functional approach*, Mouton de Gruyter, Berlin.

Dahl T. 2004, *Textual Metadiscourse in Research Articles: a Matter of National Culture or of Academic Discipline?* in "Journal of Pragmatics" 36, pp. 1807-1825.

D'Angelo L. 2010, *Creating a Framework for the Analysis of Academic Posters*, in O'Brien L.J. and Giannoni D.S. (eds), *Language Studies Working Papers 2*, pp. 38-50.
D'Angelo L. 2012, *Academic Poster Presentations: Mapping the Genre*, in M. Gotti (ed.), *Academic Identity Traits. A Corpus-Based Investigation*, Peter Lang, Bern, pp. 263-282.
D'Angelo L. 2016a, *Academic posters. A textual and visual metadiscourse analysis*, Peter Lang, Bern.
D'Angelo L. 2016b, *The academic poster genre. Friend or foe?*, in Hyland K. and Shaw P. (eds), *The Routledge Handbook of English for Academic Purposes*, Routledge, London , pp. 394-404.
De Haan F. 1999, *Evidentiality and epistemic modality: Setting boundaries*, in "Southwest Journal of Linguistics" 18, pp. 83-101.
De Haan F. 2001, *The relation between modality and evidentiality*, in Müller R. and Reis M. (eds), *Modalität und Modalverben im Deutschen*, in "Linguistische Berichte, Sonderheft" 9, pp. 201-216.
De Haan F. 2005, *Encoding speaker perspective: Evidentials*, in Frajzyngier Z. and Rood D. (eds), *Linguistic diversity and language theories*, Benjamins, Amsterdam.
De Haan F. 2012, *The Relevance of Constructions for the Interpretation of Modal Meaning: The Case of Must*, in "English Studies" 9 [6], pp. 700-728.
DeLancey S. 1997, *Mirativity: The grammatical marking of unexpected information*, in "Linguistic Typology" 1, pp. 33-52.
DeLancey S. 2001, *The mirative and evidentiality*, in "Journal of Pragmatics" 33 [3], pp. 369-382.
Dubois B.L. 1985a, *Popularization at the Highest Level: Poster Sessions at Biomedical Meetings*, in "International Journal of the Sociology of Language" 56, pp. 67-84.
Dubois B.L. 1985b, *Poster Sessions at Biomedical Meetings: Design and Presentation*, in "The ESP Journal" 4, pp. 37-48.
Fitneva S.A. 2001, *Epistemic marking and reliability judgments: Evidence form Bulgarian*, in "Journal of Pragmatics" 33 [3], pp. 401-420.
Givón T. 1982, *Evidentiality and Epistemic Space*, in "Studies in Language" 6 [1], pp. 23-49.
González M. 2005, *An approach to Catalan evidentiality*, in "Intercultural Pragmatics" 2 [4], pp. 515-540.
Gross A., Harmon J., and Reidy M. 2002, *Communicating Science: the Scientific Article from the 17th century to the present*, Oxford University Press, Oxford.
Kress G. 2010, *Multimodality. A Social Semiotic Approach to Contemporary Communication*, Routledge, London.
Halliday M.A.K. and Martin J.R. 1993, *Writing Science*, The Falmer Press, London.
Hobbs P. 2003, *The use of evidentiality in physicians' progress notes*, in "Discourse Studies" 5, pp. 451-478.
Hyland K. 2017, *Metadiscourse: What is it and where is it going?*, in "Journal of Pragmatics" 113, pp. 16-29.
Hunston S. 2002, *Corpora in Applied Linguistics*, Cambridge University Press, Cambridge.
Ifantidou E. 2001, *Evidentials and Relevance*, Benjamins, Amsterdam.
International Committee of Medical Journal Editors (ICMJE) 1991-2017, *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*. http://www.icmje.org/recommendations/ (09.03.2018).
Istituto Superiore di Sanità (National Health Institute). www.iss.it (09.03.2018).
Jakobson R. 1957, *Shifter, Verbal Categories and the Russian Verb*, in Waugh L.R. and

Halle M. (eds), 1987, *Russian and the Slavic Grammar-Studies 1931-1981*, de Gruyter, Berlin, pp. 41-58.
Kamio A. 1997, *Territory of Information*, Benjamins, Amsterdam.
Lazard G. 2001, *On the grammaticalization of evidentiality*, in "Journal of Pragmatics" 33 [3], pp. 359-367.
Maci S.M. 2011, *Genre variation in medical discourse: the case of medical posters*, in Sarangi S., Polese V. and Caliendo G. (eds), *Genre(s) on the Move. Hybridization and Discourse Change in Specialized Communication*, ESI, Napoli, pp. 169-190.
Maci S.M. 2012a, *The Genre of Medical Posters*, in Gotti M. (ed.), *Academic Identity Traits*, Peter Lang, Bern, pp. 283-301.
Maci S.M. 2012b, *Poster Makers Should Think as Much about Show Business as Science. The Case of Medical Posters in a Diachronic Perspective*, in Garzone G., Catenaccio P. and Degano C. (eds), *Genre Change in the Contemporary World*, Peter Lang, Bern, pp. 77-98.
Maci S.M. 2015a, *"If MSM are frequent testers there are more opportunities to test them": conditionals in medical posters – a corpus-based approach*, in Thompson P. and Diani G. (eds), *English for Academic Purposes. Approaches and Implications*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 127-150.
Maci S.M. 2015b, *"These data support the provocative view that…": Evaluation in medical academic posters*, in Bondi M., Cacchiani S. and Mazzi D. (ed.), *Discourse In and Through the Media. Recontextualizing and Reconceptualizing Expert Discourse*, Cambridge Scholars Publishing, Cambridge upon Tyne.
Maci S.M. 2016, *A framework for medical poster multimodal analysis: a pilot study*, in Daniele F. and Garzone G. (eds), *Communicating Science, Popularizing Science*, FrancoAngeli Editore, Milano, pp. 49-71.
McEnery T. and Wilson A. 1996, *Corpus Linguistics*, Edinburgh University Press, Edinburgh.
MacIntosh-Murray A. 2007, *Poster Presentations as a Genre in Knowledge Communication: A Case Study of Forms, Norms, and Values*, in "Science Communication" 28 [3], pp. 347-376.
Mair C. 1991, *Quantitative or qualitative corpus analysis? Infinitival complement clause in the Survey of English Usage corpus*, in Johansson J. and Stenström A.B. (eds), *English Computer Corpora: Selected papers and research guide*, Mouton de Gruyter, Berlin, pp. 67-80.
MEDLINE, www.ncbi.nlm.nih.gov/pubmed/ (08.03.2018).
Miles M.B. and Huberman A.M. 1994, *Qualitative data analysis: An expanded sourcebook*, 2nd ed., Sage, Thousand Oaks, CA.
Mitrany D. 2005, *Creating Effective Poster Presentations: The Editor's Role*, in "Science Editor" 28 [4], pp. 114-116.
Mocini R. 2015, *Evidential Devices in English Medical Journals*, in Salvi R. and Bowker J. (eds), *The Dissemination of Contemporari Knwoledge in English*, Peter Lang, Bern, pp. 73-97.
Mushin I. 2001, *Evidentiality and Epistemological Stance: Narrative Retelling*, Benjamins, Amsterdam.
Mushin I. 2013, *Making knowledge visible in discourse: Implications for the study of linguistic evidentiality*, in "Discourse Studies" 15 [5], pp. 627-645.
National Institute for Health and Care Excellence (NICE). https://www.evidence.nhs.uk/ (09.03.2018).

Nwogu N.K. 1997, *The Medical Research Paper: Structure and Functions*, in “English for Specific Purposes” 16 [2], pp. 119-138.
Nuyts J. 2001, *Subjectivity as an evidential dimension in epistemic modal expressions*, in “Journal of Pragmatics” 33 [2], pp. 383-400.
Nuyts J. 2015, *Subjectivity: Between Discourse and conceptualization*, in “Journal of Pragmatics” 86, pp. 106-110.
Palmer F.R. 1986, *Mood and Modality*, Cambridge University Press, Cambridge.
Papafragou A. 2000, *Modality: Issues in the Semantics-Pragmatics Interface*, Elsevier Science, Oxford.
Pearce E.L.F. 1992, *Proceedings, IADR Council Meeting. Committee Reports*, in “Journal of Dental Research” 71 [1], pp.1677-1681.
Plungian V.A. 2001, *The place of evidentiality within the universal Grammatical space*, in “Journal of Pragmatics” 33 [3], pp. 349-357.
Quirk R., Greenbaum S., Leech G. and Svartvik J. 1985, *A Comprehensive Grammar of the English Language*, Longman, London.
Rowe N. 2012 *Academic & scientific poster presentation: visualising knowledge as a means to facilitate knowledge transfer,* in Ilic D. (ed). *Knowledge transfer: practices, types and challenges,* Nova Science Publishing, New York, 123-140.
Rowe N. 2017, *Academic & Scientific Poster Presentation. A Modern Comprehensive Guide,* Springer, Cham.
Sackett D.L., Rosenberg W.M.C., Gray J.A.M., Haynes R.B. and Richardson W.S. 1996, *Evidence based medicine: what it is and what it isn’t*, in “British Medical Journal” 312, pp. 71-72.
Scott M. 2012, *WordSmith Tools version 6*, Lexical Analysis Software, Stroud.
Skelton J.R. and Edwards S.J. 2000, *The Function of the Discussion Section in Academic Medical Writing*, in “British Medical Journal” 320, pp. 1269-1270.
Squartini M. 2008, *Lexical vs. grammatical evidentiality in French and Italian*, in “Linguistics” 46 [5], pp. 917-947.
Squartini M. 2012, *Evidentiality in interaction: The concessive use of the Italian Future between grammar and discourse*, in “Journal of Pragmatics” 44 [15], pp. 2116-2128.
Squartini M. 2018, *Mirative extensions in Romance: Evidential or epistemic?*, in Guentchéva Z. (ed.), *Epistemic modality and evidentiality in a cross-linguistic*, Mouton de Gruyter, Berlin, pp. 196-214.
Swales J.M. 1990, *Genre Analysis. English in Academic and Research Settings*, Cambridge University Press, Cambridge.
Swales J.M. 2004, *Research Genres. Explorations and Applications,* Cambridge University Press, Cambridge.
Swales J.M. and Feak C. 2000, *English in Today’s Research Settings*, Cambridge University Press, Cambridge.
Thomas S. and Hawes T.P. 1994, *Reporting Verbs in Medical Journals*, in “English for Specific Purposes” 13 [12], pp. 129-148.
Tognini-Bonelli E. 2001, *Corpus Linguistics at Work*, John Benjamins, Amsterdam.
van der Auwera J. and Plungian V.A. 1998, *Modality’s Semantic Map*, in “Linguistic Typology” 2, pp. 79-124.
Vihla M. 1999, *Medical Writing. Modality in Focus*, Radopi, Amsterdam.
Willett T. 1988, *A cross-linguistic survey of the grammaticalization of evidentiality*, in “Studies in Language” 12, pp. 51-97.
Wilson L. 1942, *The Academic Man: A Study in the Sociology of a Profession*, Oxford University Press, New York.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 93-119
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p93

VAGUE LANGUAGE IN THE MMR VACCINE CONTROVERSY A corpus-assisted discourse analysis of its functional use[1]

ANNA FRANCA PLASTINA,[1] ROSITA MAGLIE[2]
[1]UNIVERSITÀ DELLA CALABRIA, [2] UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Abstract – This paper explores how vague language is used in multiple forms of specialised knowledge which have contributed to triggering the ongoing debate on MMR vaccine-induced autism. The controversy has been stirred up by the publication of the Wakefield et al. paper in 1998, considered one of the most serious cases of fraud in medical history. In the current paper, contending discourses shaped by different scientific and lay agents are considered in a diachronic perspective in order to investigate how the legitimacy of the knowledge claim is disputed through the functional use of approximators, vague quantifiers, epistemic stance markers, subjective stance markers and general extenders/placeholders. Accordingly, a corpus of various text types is introduced to disclose how fraudulent scientific knowledge is produced, propagated in the public domain as a medical myth, and refuted through investigative journalism which has led to the rare practice of retraction of the 1998 research article. A corpus-assisted approach to discourse analysis is adopted to unpack the functions these vague language categories play in this evolving process of knowledge production, reception and reconstruction, which allows new controversial interpretations of the same knowledge to emerge. Quantitative and qualitative findings shed light on how the set of vague categories functionally operate to cast doubts about scientific knowledge, and strengthen its assumptions on the divide between 'good' and 'bad science'. Ultimately, the study reveals how vague language can be artfully deployed as a covert persuasive technique to undermine public confidence in the benefits of vaccination, also by drawing on the use made of vagueness by the scientific community to express uncertainty as part of the ethical practice of advancing new knowledge claims.

Keywords: vague language categories; corpus-assisted discourse analysis; specialised knowledge dissemination; MMR vaccine controversy; medical fraudulence and myths.

[1] Although this research was jointly conducted by both authors, Anna Franca Plastina is responsible for sections 1, 3 and 4.2; Rosita Maglie for sections 2, 4.1 and 5.

1. Vague language and specialised knowledge

The importance of the audience-oriented feature of vague language (VL) in knowledge claim design has been approached in several studies on medical and scientific discourse, where greater attention has been given, however, to the rhetorical device of hedging over other vague categories (e.g. Dubois 1987; Myers 1989; Salager-Meyer 1994; Plastina, Del Vecchio 2014). In this respect, Channell (1994, p. 18) proposed a taxonomy which covers three main categories of vagueness, namely vague additives, including vague approximators (e.g. *about*) and tags referring to vague category identifiers (e.g. *and stuff like that*); vagueness through lexical choice and quantifiers (e.g. *thingy*; *tons of*); vagueness by implicature in context (e.g. *six feet tall* for six feet and a quarter of an inch). Hence, VL units "can be contrasted with another word or expression which appears to render the same proposition", but are "purposely and unabashedly vague" (Channell 1994, p. 20). In a broader taxonomic view, Zhang (2015, p. 36) considers VL in terms of "stretchers", or expressions with a "fluid and elastic characteristic", and distinguishes four main lexical categories, namely approximate stretchers (approximators, vague quantifiers); general stretchers (general terms), scalar stretchers (intensifiers, softners); epistemic stretchers (epistemic stance markers).

Moreover, vagueness is a pervasive feature of natural language, which is basically connected to the semantic concept of "*underdetermination* and *openness* of meaning" (Égré, Klinedinst 2011, p. 7, original emphasis). Accordingly, VL "remains vague in context, rather than becoming precise", but "does not disrupt ease of communication" (Sabet, Zhang 2015, p. 7). More importantly, VL can be considered as an indicator of intersubjectivity (Overstreet, Yule 2002), thus differing from both indirect and inexplicit language.[2] Its feature of non-specificity is strictly related to the context-dependable variable of "shared knowledge" (O'Keeffe 2003), whereby "any social group sharing interests and knowledge employs non-specificity in talking about their shared interest" (Channell 1994, p. 193). While VL thus marks in-group membership, its use is not, however, merely "a viable communicative option in most scientific writing" (Hyland 1998a, p. 256). It is rather deliberately used in this genre as a "threat-minimizing strategy" (Salager-Meyer 1994, p. 150), serving the overall functional purpose of meeting "adequacy and acceptability conditions" (Hyland 1996, p.437). The linguistic phenomenon of VL therefore appears to play a preeminent role especially when innovative medical knowledge claims are disseminated

[2] See Cheng and Warren (2003) for a detailed distinction between indirectness, inexplicitness and vagueness.

within the scientific community through the publication of research articles. In this interactional context, VL helps negotiate the status of medical knowledge claims in an adequate manner even during the review and revision processes regarding the publication of research articles (Myers 1985). Hence, authors exploit different devices to "stretch language elastically in discursive negotiations" (Zhang 2011, p. 571) for the primary purpose of soliciting acceptance of their medical knowledge claims by an engaged community of readers (Hyland 1998b). In other words, VL allows medical scientists to design their knowledge claims in ways that attenuate criticism (Zhang 2011) and find acceptance for their speculations, while also contributing to building their reputation for the advancements made in the field of medicine both within the scientific community and beyond.

This paper then explores how different VL categories may be exploited to negotiate "moment-to-moment communicative needs" (Zhang 2015, p. 55), which account for the audience-oriented aspect of knowledge claims and their dissemination across both scientific and lay communities. The research advocates the importance of exploring "the different settings in which knowledge circulates, setting out from the supposition that science forms part of the practices of human communities" (Calsamiglia, Ferrero 2003, p. 147). Focusing on the controversial claim that the measles-mumps-rubella (MMR) vaccine is linked to the condition of autism, the study is based on a corpus-assisted discourse analysis (see Baker *et al*. 2008) of a collection of sample scientific and popular texts dealing with the issue of MMR-induced autism. The aim is to seek how varying degrees and types of vagueness are conveyed through a range of linguistic units for the peculiar unethical purpose of obscuring false scientific meaning, and also to investigate how VL categories are further deployed in popular discursive reconstructions of the same debatable knowledge claim. Two main research questions are thus addressed: 1.How do vague categories operate to obscure false scientific meanings?; 2. How are they deployed across the scientific and lay discursive communities to disseminate fraudulent specialised knowledge?

Overall, the investigation attempts to highlight how vague categories are purposively used across the two communities as different vehicles of the unique discursive practice of disseminating fraudulent specialised knowledge and propagating medical misbeliefs. It may also shed light on the ways in which VL contributes to widely conditioning parents' refusal of vaccinating their children, thus leading to the growing concern of increased MMR morbidity and mortality.

2. The MMR vaccine controversy: disseminating 'bad science'

In a diachronic perspective, the worldwide controversy on the safety of the MMR vaccine dates back to the publication of the Wakefield et al. paper in the prestigious medical journal, *The Lancet* in 1998. The article presented the case of twelve children with assumed brain disorders and specifically correlated brain damage and the onset of a new bowel syndrome (autistic enterocolitis) with the fact that they had previously received the MMR vaccine. Linguistic research has focussed on the "harms of hedging" caused by Wakefield's article (see Kolodziejski 2014). However, it has largely underestimated the "discursive gaps" created by other potential vague devices, which allow for alternate interpretations as the scientific text has passed from the medical community to public contexts. The alarming medical breakthrough prompted, in fact, the publication of new medical articles by other researchers, who failed, however, to replicate Wakefield's research and rejected his subjective claims through the conventional rhetorical use of VL to mark scientific uncertainty. Nevertheless, the "medical myth" (see Vreeman, Carroll 2007) started to spread in the public arena even thanks to a video-news release distributed by the Royal Free Hospital (Wakefield's workplace) and a science-by-press conference held by Wakefield himself.

While the use of VL in these discursive practices appeared to publicly grant Wakefield the claim to expertise, the legitimacy of the disputed claim was particularly contended by the journalist Brian Deer, who started writing newspaper articles the same year Wakefield's paper was published. In his investigations for *The Sunday Times*, and the UK *Channel 4* TV network,[3] Deer highlighted how Wakefield's findings merely relied on the memories and assertions of the sample children's parents and were not evidence-based although their presumed "regressive autism" had appeared days after receiving the MMR vaccine. Moreover, Deer found that Wakefield had been secretly payrolled by a lawyer, Richard Barr, to produce evidence against the MMR vaccine two years before the publication of the *Lancet* paper and had filed a patent on a presumed "safer" single measles vaccine, nine months before the 1997 science-by-press conference during which he called for single vaccines. Another crucial issue unearthed by Deer refers to Wakefield's deliberate tampering with results: even though the hospital's clinicians and pathology service had not detected any correlation between autism and the MMR vaccine in these children, Wakefield misreported and misrepresented their histories and diagnoses, thus acting entirely in his own

[3] See https://briandeer.com/mmr/andrew-wakefield.htm.

interests. Furthermore, the British and American anti-vaccine movements were gaining strength supported by Wakefield's claim of the link between what he called an "epidemic of autism" and the MMR vaccine, which continued to be constantly spread through conferences (e.g. "Defeat Autism Now" conference, 1998) and the media (e.g. the CBS network's 60 Minutes programme, November 2000). In parallel, Deer's investigative outcomes started to circulate extensively to the point that Wakefield's *Lancet* paper was retracted in 2004, and a professional misconduct hearing was held by the UK General Medical Council (GMC) between 2004-2010. Notwithstanding the burden of proof on his misconduct, Wakefield issued a claim for libel following the broadcast of Deer's "MMR-What they didn't tell you" in November 2004. While in 2007 Wakefield agreed to abandon his claim and to pay the defendants' costs, in 2010 the GMC found him guilty of serious professional misconduct on a number of charges and subsequently erased him from the UK doctors' register. *The Lancet* fully retracted Wakefield's paper the same year and in 2011 *The British Medical Journal* (BMJ) published three major reports by Deer on the MMR fraud. As the first investigative journalist to publish in a medical journal, Deer was named the UK specialist reporter of the year by the British Press Awards. Deer's collection of overwhelming evidence and the institutional system of retractions have alerted the scientific community, whereby "the publication of false science […] cause[d] the dramatic loss of reputation for the individual scientist associated with the falsification" (Furman *et al*. 2012, p. 278). Paradoxically, however, the "medical myth" and the anti-MMR campaign still continue to be in the spotlight; the media appear to persist in artfully taking full advantage of VL as a covert persuasive technique. Based on the premise of the evolving process of MMR knowledge production, reception and reconstruction in multiple scientific and popular text types, the present investigation thus seeks to uncover how VL contributes to constructing "bad science" and reconstructing it as "good science" in popular contexts.

3. Corpus and methodology

The corpus is composed of 64 texts (117,709 words) selected from scientific and popular web-based sources according to the criterion of different text types in which vague categories may be used to mediate the pros and cons of the MMR vaccine controversy. For present purposes, the collection was divided into two sub-corpora, namely the Scientific MMR (SMMR) subcorpus and the Popular MMR (PMMR) one. The SMMR is made up of a total of 39 texts counting 65,293 words and covers seven different text types including scientific abstracts, Brian Deer's BMJ articles,[4] Wakefield's journal articles, the *Lancet* journal retractions, scientific journal correspondence, editorials and Wakefield's science-by-press conference;[5] the PMMR is made up of 25 texts amounting to 52,416 words, and includes seven different text types, namely a blog, press interviews, radio shows, TV interviews, internet videos, TV news and newspaper articles. The contribution of fourteen different text types is here seen as helping increase balanced corpus representativeness. A combination of quantitative and qualitative methods has been adopted in the study as "mutually supportive methodologies" (Mautner 2009), thus combining a corpus linguistic approach and critical discourse analysis for in-depth analysis and data reliability (Partington 2006). The analysis was first conducted through the basic corpus linguistic methods of frequency counts and concordance lines using Wordsmith 5.0 concordancer to identify the occurrence of vague categories in the two subcorpora, based on Channell (1994) and Zhang's (2015) taxonomies. A limited use was thus made of these corpus-based techniques to provide preliminary findings before the qualitative analysis (Baker *et al.* 2008). This was conducted to bring to light the different vague categories deployed in the representation of specialised knowledge as "bad science" and in its popular redefinition as apparently "good science". In this second step of the analysis, data was coded according to the following five categories: 1. *approximators* (smaller or bigger than the exemplar number, bigger than the exemplar number or smaller than the exemplar number); 2. *vague quantifiers* (multal, paucal, negative or neutral), whereby multal quantifiers (e.g. *many*) are intensifiers in assertive contexts, paucals (e.g. *a bit of*) are downtoners in both assertive and negative contexts, negative quantifiers (e.g. *few*) are minimisers in negative contexts, and neutral ones (*several*) are used in both contexts (Zhang 2015, p. 29); 3. *epistemic stance markers* (doubt, actuality and reality, source of knowledge, limitation of knowledge, viewpoint and perspective); 4. *subjective stance markers* (speaker's attitude toward proposition, style and manner of speaking, or imprecision/hedging); 5.*general extenders* (adjunctive or disjunctive) and *placeholders* (dummy nouns). Although these vague categories were treated together for their common pragmatic interpersonal function of sharing MMR knowledge, general extenders were understood to "have nonspecific reference or 'general' reference, and […] 'extend' otherwise grammatically complete utterances" (Overstreet, Yule 1997, p. 251) usually in clause-final position; placeholders were considered as

[4] As a scientific journalist, Deer was exceptionally asked to write three articles for the BMJ. Thus, these articles appear to belong to a hybrid genre compared to Wakefield's ones.

[5] The practice by which scientists put unusual focus on promoting their questionable research findings by turning to the media when they are unlikely to gain consensus within the scientific community (Jerome 1989).

"dummy nouns which stand for item names" (Channell 1994, p. 164). Moreover, both *adjunctive general extenders* (e.g. *and everything*) and placeholders (e.g. *something*) serve the primary function of signaling that "'there is more' to infer, thus marking assumed reciprocity of perspectives" (Overstreet 2005, p. 1855); disjunctive general extenders, instead, are "tied to indicating potential alternatives, and hence hedging on what has been said" (Overstreet 2005, p. 1855) for different functional purposes. A critical discourse analytical approach was finally adopted for interpretation and explanation of all these vague categories within their social context of use (Fairclough 2013).

4. Analysis and results

4.1. The scientific sub-corpus

4.1.1. Approximators and vague quantifiers

Approximators are not particularly used across the different subcorpus text types (*N*=42), thus suggesting that their function in scientific discourse is quite irrelevant, as shown in Table 1.

	Approximators		
Scientific Genre	*Smaller or bigger than the exemplar number* (*N*=16)	*Bigger than the exemplar number* (*N*=10)	*Smaller than the exemplar number* (*N*=16)
Abstracts (11)	Or (5) About (1) Approximately (1)	Over (2)	Almost (1) Nearly (1)
Deer's articles (21)	About (4) Or (3)	At least (3) Or so (1) Over (1)	Almost (5) Nearly (4)
Wakefield et al. papers (7)	Around (1) Or (1)	At least (1) Over (1)	Less than (2) Almost (1)
Correspondence (2)	=	=	Almost (2)
BMJ Editorial (1)	=	At least (1)	=
Lancet retractions (0)	=	=	=
Wakefield's Science-by-Press conference (0)	=	=	=

Table 1
Approximators per scientific genres in the SMMR subcorpus.

In particular, the limited use of numerical approximators appears to point to the expected scientific practice of providing exact numbers, whereby only Brian Deer, as a journalist and not a scientist, is consistently found to make a more frequent use of this category in his *BMJ* articles (*N*=21). The use of *less than* (2) in the abstract and result sections of two Wakefield et al. papers appears, instead, to be an acceptable practice in medical research articles as exact numbers can be found in other parts of the papers, as well as in the figures and tables. A*round*, *at least*, and *or* all occur only once in *The Lancet* retracted paper to approximate age, year, allele, thus assuming that there is no fixed pattern when a biological process, a disease onset or a genetic inheritance take place in human beings, as indicated respectively in Examples (1)-(3):

(1) Vitamin B12 is essential for myelinogenesis in the developing central nervous system, a process that is not complete until **around** the age of 10 years (*Wakefield et al., Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children, The Lancet, 1998)*
(2) Disintegrative psychosis is typically described as occurring in children after **at least** 2-3 years of apparently normal development (*Wakefield et al., Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children, The Lancet, 1998*).
(3) individuals inheriting one **or** two C4B null alleles may not handle certain viruses appropriately (*Wakefield et al., Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children, The Lancet, 1998*).

The more recurring use of *or* (*N*=5) in the abstracts mainly suggests that exact MMR vaccine dosage is not a factual determinant of *pervasive developmental disorder/autism*, as indicated in Example (4):

(4) Thus, <u>no relationship was found</u> between pervasive developmental disorder rates/autism and 1- **or** 2-dose measles-mumps-rubella immunization schedule (*Fombonne et al., Pervasive Developmental Disorders in Montreal, Quebec, Canada: Prevalence and Links With Immunizations, Pediatrics, July 2006*).

There is uniformity of context of occurrence when Deer uses *almost*, *nearly* and *about* in his BMJ articles to respectively approximate the impact of fraud research in terms of the *Lancet* children's health state (Example 5a), and the economic drive behind it (Example 5b):

(5) a. In **almost** all the children, […] the hospital's pathology service found the children's colons to be largely normal, but a medical school "review" <u>changed</u> the results […] And as an alternative explanation for any inflammation that was present, **nearly** all of the children had constipation with megarectum (<u>unreported</u> in the paper), which specialists say can cause cellular changes (*Brian Deer, Wakefield's "autistic enterocolitis" under the microscope, BMJ, 15 April 2010*).

b. It is estimated that by year 3, income from this testing could be **about** £3,300,000 rising to **about** £28,000,000 […] (*Brian Deer*, *How the vaccine crisis was meant to make money, BMJ, 11 January 2011*).

While in (a), the use of *smaller* approximators than the number of sample children highlights the inexactness of medical test results (*changed*, *unreported*), the calculation in (b) oscillates between a large numerical interval (from smaller to bigger), which denotes the *estimated* enormous profit Wakefield expected to gain from the study.

As for vague quantifiers, *multal* and *neutral* subtypes are the most frequent ones in the subcorpus (*N*=156) with a more meaningful occurrence of *some* in both Wakefield and Deer's works (see Table 2). In the 1998 article, Wakefield claimed the discovery of a *new* syndrome, although the exact number is used with a *neutral* value (*some cases, some children*) in order to attenuate possible criticism (Zhang 2011), as shown in Example (6a). Drawing on the importance of medical breakthroughs (*it is important to consider*), the subsequent occurrence of *some* seems to resort to its more conventional scientific function of expressing tentativeness (Hyland 1996), also strengthened by the modality of possibility (*might potentially*). This deliberate use helps build the credibility of the initial claim *soon after MMR vaccine*.

(6) a. a new syndrome has been reported in children with autism […] (autistic enterocolitis), in **some** cases *soon after MMR vaccine*. […] it is important to consider why **some** children <u>might potentially</u> react aberrantly to a vaccine when the <u>majority</u> do not (*Wakefield, Enterocolitis, autism and measles virus, Mol Psychiatry 7 Suppl., 2002*).

b. **Some** children were reported to have experienced first behavioural symptoms *within days of MMR*, but the records documented these as starting **some** months after vaccination […] there is a suggestion that **some** of his problems [child 4] may have started *before vaccination* […] **some** of the authors met and agreed that the paper, already intended for submission to the high impact journal, was accurate […] (*Brian Deer, Pathology reports solve "new bowel disease" riddle, BMJ, 9 November 2011*).

In Example (b), instead, vague neutrality is used by Deer to objectively highlight the arguable correlation between behavioural symptoms and their temporal onset, grounded in evidential proof (*records documented*), which point to *some months later*, and *before vaccination*. In spite of data manipulation, Deer further quantifies the authors as *some* to implicitly indicate the unimportance of the precise number (but at least two *met and agreed*) compared to the scandalous fact that some consent was reached on data accuracy.

	Vague Quantifiers			
Scientific Genre	*Multal* (*N*=61)	*Paucal* (*N*=1)	*Negative* (*N*=3)	*Neutral* (*N*=95)
Abstracts (6)	Number of (3)	=	=	Several (2) Some (1)
Deer's articles (59)	Many (11) Much (6) A lot of (1) A Load of (1) A Number of (1)	A bit (1)	Little (1) Few (1)	Some (28) A few (5) A couple of (2) Several (1)
Wakefield et al. papers (67)	Many (13) Numbers of (7) Much (3)	=	=	Some (37) Several (6) A few (1)
Correspondence (9)	A number of (2) Many (1)	=	=	Some (6)
BMJ Editorials (6)	Many (2) A great deal of (1) A number of (1)	=	Few (1)	Several (1)
Lancet Retractions (1)	Much (1)	=	=	=
Wakefield's Science-by-Press Conference (12)	Many (3) Much (1) A lot of (1) A great deal of (1) A number of (1)	=	=	Some (3) Several (1) Couple (1)

Table 2
Vague quantifiers per scientific genres in the SMMR subcorpus.

4.1.2. Stance markers

In five out of the seven text types analysed, epistemic stance markers were found to mostly signal *doubt* (*N*=50) and point to *actuality/reality* (*N*=19), while in four types they were also used to indicate *source of knowledge* (*N*= 27). The most recurring markers were the modal *may* + *be* in Wakefield et al. papers and the verb *I think* in Wakefield's science-by-press conferences and in Deer's BMJ articles, as shown in Table 3.

	Epistemic Stance Markers				
Scientific genre	*Doubt* (*N*=50)	*Actuality and Reality* (*N*=19)	*Source of Knowledge* (*N*=27)	*Limitation* (*N*=7)	*Viewpoint/ Perspective* (*N*=6)
Abstracts (3)	=	=	=	Typically (3)	=
Deer's articles (53)	(I, you etc.) think (11) Maybe (7) Probably (1)	Really (6) Actually (5) In fact (4)	According to (7) Apparently (5) Reputedly (2) Evidently (1)	Typically (1)	In (my, your etc.) view (3)
Wakefield's et al. papers (37)	May be (21) Probably (1) Most Likely (1)	Truly (1)	Apparently (5) According to (5)	Typically (2) Mainly (1)	=
Correspondence (7)	Maybe (1)	In fact (2)	According to (1)	=	In (my, your etc.) opinion (3)
BMJ Editorials (2)	Perhaps (1)	In fact (1)	=	=	=
Lancet Retractions (1)	=	=	According to (1)	=	=
Wakefield's Science-by-Press Conference (6)	Perhaps (2) I think (4)	=	=	=	=

Table 3
Epistemic stance markers per scientific genres in the SMMR subcorpus.

The modal *may* not only indicates that there is a possibility that a medical phenomenon is plausible (*majority* vs. *subgroup* of children) (7a), or is true in some circumstances (*after measles, mumps and rubella immunisation*) (7b), but it also signals the authors' presence in a text. As observed by Halliday and Matthiessen (2004, p.624), "[m]odality represents the speaker's angle", as it indicates his/her knowledge and commitment to the truth of the proposition (Palmer 2001). Wakefield et al. appear to be uncertain (7a) about the correlation between chronic enterocolitis and MMR immunisation (7b), but their mention only of the MMR vaccine discourages alternative interpretations, thus leading the reader to share their own views.

(7) a. Whether such abnormalities may be seen in the majority of children with autism, or are restricted to a subgroup with clear regression, remains uncertain (*Torrente et al. (Wakefield included), Small intestinal enteropathy with epithelial IgG and complement deposition in children with regressive autism, Molecular Psychiatry, 2002*).

b. a chronic enterocolitis in children that maybe related to neuropsychiatric dysfunction. In most cases, onset of symptoms was after measles, mumps, and rubella immunisation. Further investigations are needed to examine this syndrome and its possible relation to this vaccine (*Wakefield et al., Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children, The Lancet, 1998*).

While in the medical texts the author's presence tends to be minimal, in the science-by-press conference Andrew Wakefield marks his presence more explicitly using the possessive adjective *my* and the first person pronoun *I*. He uses *I think* as a way of being polite to make his statement sound less forceful but, bridging science (he as a health scientist) with conscience (he as a human being whose *mind* tells him to do what it is right), he induces the audience to strongly believe he is acting in the interests of patients' safety (8).

(8) There is sufficient anxiety in my own mind for the long term safety of the polyvalent vaccine—that is, the MMR vaccination in combination—that I think it should be suspended in favour of the single vaccines (*Wakefield, science-by-press conference, 26 February 1998*).

Instead, he is actually taking advantage of parents' state of vulnerability and of their anxious search (*looking for answers*) in the hope that their youngsters may recover (9). Under these conditions, they are thus more likely to place complete trust in what they are told by doctors. This explains Deer's detached commitment *(I do not think*) to the remote possibility of a *connection with MMR* (*only suspicion*) based on the family's uncertainty.

(9) As for a connection with MMR, there was only suspicion. I do not think his family was sure, one way or the other. When I asked why they took him to the Royal Free, his father replied: "We were just vulnerable, we were looking for answers" (*Brian Deer, How the case against the MMR vaccine was fixed, BMJ, 5 January 2011*).

Furthermore, subjective stance markers were the least frequent category found in the SMMR subcorpus with only 13 instantiations occurring in Deer's articles, Wakefield et al. papers, and correspondence (see Table 4). The feasible difficulty of explaining a medical phenomenon lies at the basis of the use of two contrasting subjective stance markers: *surprisingly* in Wakefield et al. papers and *curiously* in Deer's *BMJ* articles. The former is used to indicate the current limitation of medical knowledge without providing any evidence whatsoever of the claim that *upper gastrointestinal pathology, […], is also present in these children at a surprisingly high rate* (*Wakefield, Enterocolitis, autism and measles virus, Mol Psychiatry 7 Suppl., 2002)*. The objective of this attitude is only to cast doubt and undermine the reader's prior certainty; the latter marker is ironically used to illuminate the reader that *curiously, however, Wakefield had already identified such a syndrome before the project which would reputedly discover it* (*Brian Deer, How the case against the MMR vaccine was fixed BMJ, 5*

January 2011). The subjective stance marker is here backed up by irrefutable evidence that Wakefield made his assumption as far back as two years before the publication of the *Lancet* paper when he was secretly payrolled to find scientific proof against MMR vaccine.

	Subjective Stance Markers		
Scientific genre	*Speaker's Attitude toward Proposition* (*N*=2)	*Style Stance Marker (manner of speaking)* (*N*=8)	*Imprecision/ Hedging* (*N*=3)
Abstracts (0)	=	=	=
Deer's articles (8)	Curiously (1)	Simply (4)	Kind of (3)
Wakefield et al. papers (3)	Surprisingly (1)	Simply (1) Briefly (1)	=
Correspondence (2)	=	Simply (2)	=
BMJ editorials (0)	=	=	=
Lancet retractions (0)	=	=	=
Wakefield's Science-by-Press Conference (0)	=	=	=

Table 4
Subjective stance markers per scientific genres in the SMMR subcorpus.

4.1.3. Placeholders for sharing MMR knowledge

While no general extenders were recorded in the SMMR subcorpus, a small number of placeholders (*N*=27) was found to be used to mark that there could be more to say. The most frequently occurring placeholder turns out to be *something* in Wakefield's science-by-press conference and in Deer's articles (see Table 5). Although in both cases, there appears to be more MMR knowledge to infer, the placeholder is used for contrasting purposes, as suggested in Example 10 (a) and (b), respectively by Wakefield and Deer:

(10) a. **something** that we started after parents put us up to this back in 1998 and it was remarkable the benefit to children (*Wakefield, science-by-press conference, 26 February 1998*).

b. Did the scientific community ever really believe that 12 families had turned up consecutively at one hospital, with no reputation for developmental disorders, and make the same highly specific allegations – with a time-link of just days – and that there was not **something** fishy going on? (*Brian Deer, Piltdown medicine: the missing link between MMR and autism, BMJ online, 5 January 2011*).

In (a), Wakefield uses the placeholder *something* as a dummy noun, taking for granted that his audience shares knowledge about his research. On this common ground, he attempts to establish mutual agreement with his public by drawing on his deontological duty as a medical scientist and head of a research team, who is called to fulfil parents' impelling needs (*parents put us up to this*). Thus, the context-dependent use of the placeholder appears to help Wakefield gain public consensus, especially as the resulting *benefit to children* was *remarkable*. Yet, Deer in (b) questions Wakefield's utterance by using the placeholder *something* to signal a perceived "breach of the reciprocity of perspectives" (Overstreet 1999, p. 74), which is then overtly further marked by the negative qualifier *fishy*. The overall effect is to arouse his readers' suspicion of Wakefield's conduct against the scientific community's skepticism (*ever really believe*).

Scientific genre	*Placeholders (N=27)*
Abstracts (0)	=
Deer's articles (20)	Something (5) Anything (4) Things (4) Anybody (3) Somebody (2) Someone (1) Stuff (1)
Wakefield's et al. papers (0)	=
Correspondence (1)	Anything (1)
BMJ Editorials (1)	Anything (1)
Lancet Retractions (0)	=
Wakefield's Science-by-Press Conference (5)	Something (3) Someone (2)

Table 5
Placeholders per scientific genres in the SMMR subcorpus.

4.2. The popular sub-corpus

4.2.1. Approximators and vague quantifiers

A total of 56 vague approximators were found in the PMMR subcorpus (Table 6) with major occurrences in the newspaper article (*N*=16), the press interview (*N*=14), and the TV interview (*N*=12), thus suggesting how these categories seem to support the "audience-oriented" feature in these popular genres. In particular, the most frequent sub-type of approximator recorded was *about* (*N*=27), which was mainly used to vaguely indicate exemplar numbers without a precise cut-off point (*smaller or bigger than the exemplar number*).

	Approximators		
Media genre	*Smaller or bigger than the exemplar number (N=29)*	*Bigger than the exemplar number (N=16)*	*Smaller than the exemplar number (N=11)*
Blog (0)	=	=	=
Press Interview (14)	About (10)	Over (1) Or so (1) At least (2)	=
Radio Show (7)	About (5) Approximately (1)	At least (1)	=
TV Interview (12)	About (5)	At least (4) Over (1)	Less than (2)
Internet Video (6)	Roughly (1)	Over (2) At least (2)	Nearly (1)
TV News (1)	=	=	Almost (1)
Newspaper Articles (16)	About (7)	At least (2)	Nearly (5) Almost (2)

Table 6
Approximators per media genres in the PMMR subcorpus.

The effectiveness of *about* in shaping meaning making can be seen in Example (11), where Wakefield approximates the number of his publications for the subtle functional purpose of impressing the public audience so as to gain recognition for his expertise, and thus indirectly, for his knowledge claim:

(11) I have now published **about a hundred thirty, one hundred forty** peer-reviewed papers looking at the mechanism and cause of inflammatory bowel disease and then of course lately, looking at how the brain and the bowel interact in the context of children with developmental disorders such as autism (*Wakefield's press interview with the alternative medicine proponent, Dr.Mercola, April 10, 2010*).

In other words, vagueness is more salient in terms of the size of the two numbers representing the interval of approximation (*a hundred thirty*, *one hundred forty*), and the nature of the item modified (*peer-reviewed papers*).

However, when the same approximator is employed to disseminate valid scientific knowledge in the public arena, its functional use of indicating scientific uncertainty appears to be re-established, as shown in Example (12):

(12) Larry King: What causes autism, in your opinion? If not vaccines, what does? Dr. M. Wiznitzer: Well, we know that in **about 10 percent to 15 percent** of the cases, we can identify a genetic causation… (*CNN interview, April 3, 2009*).

Despite the American TV host's question seems to solicit a subjective answer (*in your opinion*), the pediatric neurologist uses the approximator *about* to convey scientific knowledge on objective grounds, whereby the vague unit is understood to modify the exemplar percentages based on the current limitations of *the cases* known. This then allows him to gain popular credibility so that the scientific community claim (*we can identify genetic causation*) is perceived as more trustworthy.

On the other hand, 333 instantiations of vague quantifiers referring to non-numerical quantities were found in the subcorpus, thus outweighing the occurrence of approximators especially in the press interviews (*N*=76), the newspaper articles (*N*=75), and the TV interviews (*N*=66). Vague quantifiers were mainly exemplified through the use of the *multal* subtype (*N*=166) and the *neutral* subtype (*N*=122) (see Table 7).

Interestingly, the major openness of meaning which these vague quantifiers appear to create is particularly reflected in the use of the multal subtype *many* (*N*=82). As an intensifier, *many* is understood to serve two diachronic functional purposes. First, it allows Wakefield to build an assertive context in which his MMR claim is boosted without producing 'true' knowledge and offering any further precise information other than that based on parents' subjective perceptions (Example 13a); secondly, it helps influential "popularisers of science", such as the English actress Jeni Barnett, to "engage in the mystification of science" (Bauer 1998, p. 79) and perpetrate deceptive information about the risk of the MMR vaccine even after a decade or so (Example 13b):

(13) a. […] in **many cases** <u>they</u> [mothers] claimed it was the measles, mumps and rubella vaccine, and then had lost their … all their acquired skills, such as speech, language, developmental milestones […] We've been looking for the cause of autism in the brain for **many, many years** and yet have failed to define it (*Twenty Twenty Television interview, February 4 1998).*

b. I, however, have talked to **many** <u>people</u> over the years – 22 years I've lived with my daughter – and over the years **many many** <u>people</u> have said the same thing, that when we were little, chicken pox, you took your kid to get the chickenpox, you made sure your child was near somebody who had it (*Jeni Barnett LBC Radio Show*, *7 January 2009*).

	Vague Quantifiers			
Media genre	*Multal (N=166)*	*Paucal (N=22)*	*Negative (N=23)*	*Neutral (N=122)*
Blog (2)	Most (1)	=	=	Some (1)
Press Interview (76)	Many (14) Much (6) A lot of (6) A number of (4) A great deal of (3) Loads of (1) Numbers of (1)	A bit (4) A little (1)	Few (2)	Some (28) Several (2) A few (3) A couple of (1)
Radio Show (41)	Many (6) Much (4) A lot of (4) Lots of (3)	A (little) bit of (3)	Little (6) Few (3)	Some (10) Several (1) A couple of (1)
TV Interview (66)	Many (25) Much (12) Lots of (3) A lot of (8) A number of (1)	A little (3) A bit (1)	Little (1) Few (1)	Some (11)
Internet Video (55)	Many (19) Much (3) A lot of (3) A * deal of (2) A number of (1)	A little (2) A bit (1)	Few (1)	Some (17) A few (3) Several (3)
TV News (18)	A lot of (3) A number of (2) Many (2) Much (2)	A little (2) A * bit (1)	=	Some (5) Couple (1)
Newspaper Articles (75)	Many (16) Much (7) A lot of (1) Numbers of (1) A * deal (1) A number of (1)	A little bit of (1) A little (3)	Little (5) Few (4)	Some (25) Several (6) A few (3) Couple (1)

Table 7
Vague quantifiers per media genres in the PMMR subcorpus.

The *neutral* subtype *some* as the other most frequent vague quantifier (*N*=97) was used in a similar vein as illustrated in Example (14), where Dr. Healey, health editor of *U.S. News and World Report* claims:

(14) We are all pro-vaccine… but there are **some** vaccines here – let's forget about autism - there are **some** *vaccines* here that one - a parent - can legitimately question *(Internet video, April 3, 2009).*

The seemingly loaded word here is *vaccines*, and *some* is used as a neutral quantifier to help the speaker avoid his commitment to the misleading issue of vaccine safety. In other words, the neutral quantificational term *some* deliberately helps overcome the more precise contradiction that there must exist at least one questionable vaccine, whereby the claim would overtly result as being false. Hence, this suggests that the numerous neutral vague quantifiers in the subcorpus contribute to creating power relations in the media social situations where scientific knowledge is misleadingly popularised, thus “producing social wrongs” (Fairclough 2013, p. 8).

4.2.2. Popular stance markers

A total of 310 epistemic stance markers were reported in the PMMR subcorpus (see Table 8) with higher occurrences of the markers *I think* (*N*=84) and *really* (*N*=83).

A predominantly ‘deliberate’ use was made of *I think* as an interactional source, whereby it appeared to mainly function as a ‘boosting’ device of propositional content and to mark a greater degree of the speaker’s personal involvement (Aijmer 1997), as shown in Example (15a):

(15) a. Jeni Barnett: So, **I think** they must have a certain amount of <u>natural immunity</u> - and I’m far, far happier for them to have developed that ‘<u>natural immunity</u>’ - than to be constantly filled with artificial substances.

b. Amanda: Then my son was born and he reacted very badly to what **I think** then was <u>the double or triple jab</u> (*Jeni Barnett LBC Radio Show, 7 January 2009*).

The epistemic marker allows the English actress to display her personal “epistemic access” to knowledge (K+), indicate her relative “epistemic rights” to know and claim authority of knowledge, besides marking her “epistemic responsibility” in terms of experiential knowledge (*natural immunity*) as one “type of knowables” (Stivers *et al*. 2011, pp. 10-17). As an upgrading marker of certainty of children’s innate immunity (*they must have*), *I think* is deliberately placed in the initial position to orient the outcome of the negotiation process with her direct interlocutor, who is expected to respond in a way that will, perhaps, change her epistemic status from K- to K+, i.e., from uncertainty to certainty of propositional content. In Example (15b), in fact, mutual agreement is reached (*reacted very badly*) on *the double or triple jab*. Also due to her popularity, the English actress is thus treated as knowing (Stivers *et al.* 2011) in spite of the fact that she further grounds her claim in an unreliable knowledge resource. The vague quantifier *a certain amount*, in fact, persuasively functions as an additional upgrading marker of *natural immunity* in order to weaken the MMR vaccine, vaguely represented through the negative connotation of *artificial substances*.

	Epistemic Stance Markers				
Media genre	*Doubt (N=132)*	*Actuality And Reality (N= 141)*	*Source of Knowledge (N=19)*	*Limitation (N=4)*	*Viewpoint/ Perspective (N=14)*
Blog (0)	=	=	=	=	=
Press Interview (120)	I think (24) Probably (5) I guess (3) Perhaps (2) Maybe (1)	Really (52) Actually (11) In fact (10) Truly (2)	According to (3)	Typically (1)	In (my, your etc.) opinion (6)
Radio Show (45)	I think (28) Probably (2) Maybe (2)	Really (10) Actually (3)	=	=	=
TV Interview (62)	I think (25) Maybe (5) Perhaps (4) Probably (2)	Really (10) Actually (7) In fact (3) Truly (1)	=	=	From * point of view (1) In (my, your etc.) opinion (4)
Internet Video (44)	Probably (1) Perhaps (2) Maybe (3) I think (5)	Really (7) In fact (6) Actually (6)	According to (8) Apparently (2)	Typically (2)	In (my, your, etc.) opinion (2)
TV News (14)	I think (7) Probably (1)	Actually (4) Really (1)	=	Typically (1)	=
Newspaper Articles (25)	Perhaps (3) Maybe (1) (I, you, etc.) think (6)	Really (3) Actually (3) In fact (2)	Apparently (1) According to (5)	=	In * view (1)

Table 8
Epistemic stance markers per media genres in the PMMR subcorpus.

When *I think* is juxtaposed with the equally frequent use of *really*, it is possible to note how contending discourses on the legitimacy of the medical myth arise, as reported in Example (16):

(16) To be **really** accepted, novel scientific findings must be repeated by others, and relevant experts are consulted on plausibility. (*Brian Deer, Doctoring the evidence: what the science establishment doesn't want you to know The Sunday Times, August 12 2012*).

Here, the investigative journalist Brian Deer uses the adverbial stance marker *really* to express his judgment of truth, based on the evidence presupposed by the reality of the scientific community regarding *novel scientific findings*. Thus, the evidential marker *really* allows Deer to evaluate the validity of the information on the basis of its evidential source, and to express his positive judgment about the factual truth of the proposition (see Palmer 2001).

In sum, while the evidential marker *really* implicitly points to Wakefield's bogus data first exposed by Deer and helps regain confidence in the rigour of scientific findings, the epistemic marker *I think* denotes how Barnett is not held accountable to the same degree for her knowledge, and serves to cast doubts on vaccine safety. Consistently with this negative view, more than half of the subjective stance markers (*N*=74) functioned as markers of imprecision/hedging (*N*=42) with a more significant occurrence of the two expressions *sort of* (*N*=23) and *kind of* (*N*=19) (see Table 9).

	Subjective Stance Markers		
Media genre	*Speaker's Attitude toward Proposition (N=9)*	*Style Stance Marker (manner of speaking) (N=23)*	*Imprecision/ Hedging (N=42)*
Blog (14)	=	=	Sort of (9) Kind of (5)
Press Interview (22)	Hopefully (4) Unfortunately (1)	Literally (3) Simply (2) Briefly (1)	Sort of (7) Kind of (4)
Radio Show (2)	=	=	Sort of (1) Kind of (1)
TV Interview (11)	=	Honestly (2) Sincerely (1) Frankly (1) Simply (2)	Kind of (4) Sort of (1)
Internet Video (13)	Unfortunately (1) Curiously (1)	Simply (5) Literally (1)	Sort of (4) Kind of (1)
TV News (3)	=	Frankly (1)	Sort of (1) Kind of (1)
Newspaper Articles (9)	Unfortunately (1) Hopefully (1)	Simply (2) Briefly (2)	Kind of (3)

Table 9
Subjective stance markers per media genres in the PMMR subcorpus.

In detail, the expression *sort of* is mainly introduced to strategically downplay scientific research by blurring its boundaries and making it seem less certain. This epistemic use was mostly employed to discredit scientific evidence against a link between vaccines and autism, as shown in Example (17):

(17) yet we don't have **any sort of** research to understand the potential risk of all those vaccines at once! So when someone tries to tell me that MMR alone doesn't cause autism but I take my child in for a vaccine appointment and they're getting six shots in 10 minutes, how am I supposed to feel reassured (*J.B. Handley defends Andrew Wakefield, YouTube, 6 January 2011*).

As parent of an autistic child and anti-vaccine activist, Handley here strategically uses *any sort of* to trace the somewhat fuzzy boundaries of MMR research. This intentionally helps draw a strong connection between the apparent lack of scientific knowledge and the "epistemic emotion" of the fear of the unknown (*how am I supposed to feel reassured*). The fact that fear is directed at the claimed absence of knowledge contributes to building the subjective epistemic value that MMR alone causes autism. Thus, fear appears to be a disvalue added to the prospect of the unknown *potential risk* primarily caused by undefined research. Ultimately, the subjective stance marker acts as a trigger to significantly lower the expected desirability of receiving the MMR vaccine as the fear of its uncertain risks appears to be higher than its unstated benefits. Yet, Handley's claim is demonstrably false as vaccine safety is closely monitored.[6]

Nevertheless, Handley's blog has gained extreme popularity until the social media company Medium.com recently decided to suspend his account in accordance with its policy of avoiding the dramatic spread of pseudoscientific claims. This cultural practice was particularly backed up by Handley's followers in their posts, where the subjective stance marker *kind of* was found to co-occur with negative expressions, as shown in Example (18):

(18) Ian: I have had a personal tragedy: my daughter has autism. I know what caused it: vaccines. **No kind of** evidence could ever persuade me otherwise. Therefore, there is no point in continuing the discussion (*Handley's blog, 22 April 2009).*

The co-occurrence of *no+kind of* points to the negative bias in the epistemic function of the stance marker. The *ad hoc* description of the blogger's *personal tragedy* serves to create the context where negativity is first introduced. Epistemic control is then covertly exercised through the deliberate use of the factive verb *I know*, which intrinsically presupposes that knowledge is constructed as the result of prior sensory experiences, or as "the basis of 'evidentials' ('I know because I see' […])" (Wierzbicka 1996, p. 49). Hence, the situational connection with negativity offers fertile ground for the strategic use of hedging, whereby *kind of* appears to be used to offset a negative reaction to scientific evidence, thus contributing to "socially shaping" (Fairclough 2013, p. 92) a negative interpretation by other bloggers.

4.2.3. General extenders of MMR knowledge

Disjunctive general extenders (N=70) significantly outweighed adjunctive ones (N=28) with a predominant occurrence of *or something* (49) followed by the adjunctive *and things like that* (26) (Table 10). No placeholders were, instead, recorded. This suggests that general extenders were mostly used in the PMMR subcorpus for the functional purpose of downgrading information, especially about people who do not entirely support the MMR-autism claim.

[6] See, for example, the numerous published studies on vaccine safety on the Centers for Disease Control website (www.cdc.gov/vaccinesafety/research/publications).

	General Extenders	
Media genre	*Adjunctive (N=28)*	*Disjunctive (N=70)*
Blog (2)	and things like that (1)	or whatever (1)
Press Interview (31)	and things like that (3)	or something (23) or anybody else (5)
Radio Show (21)	and things like that (8) and stuff like that (1)	or something (9) or anything (3)
TV Interview (10)	and things like that (4) and stuff like that (1)	or something (3) or anything (2)
Internet Video (20)	and things like that (5)	or something (7) or anything else (7) or somewhere (1)
TV News (2)	and things like that (2)	=
Newspaper Articles (12)	and things like that (3)	or something (7) or anybody (2)

Table 10
General extenders per media genres in the PMMR subcorpus.

As a case in point, Example (19) shows how Wakefield creates evasive meaning through the use of *or something*, further reinforced by the other disjunctive extender *or anything else*. The overall purpose is to publicly belittle the figure of a senior British health official, acting as a whistleblower in apparent good faith in Wakefield's concern about MMR vaccine safety with the manipulative effect of delighting his audience, as reported by Deer.[7]

(19) He described himself as a whistleblower **or something** … he had rung on two occasions previously… he would not leave his name **or anything else** and introduced himself as George *(Internet video, Wakefield, The Whistleblower, 25 August 2014).*

Thus, the disjunctive general extender seems to represent a powerful device, which is subtly employed to share apparently confidential information with the audience on emotional grounds, despite the whistleblower's story was already well known.[8]

On the other hand, the adjunctive general extender *and things like that* was mostly introduced to recall shared knowledge, and thereby extend biased views to include more objective scientific facts, as in Example (20):

[7] See http://briandeer.com/solved/whistleblower-betrayed.htm.

[8] According to Deer, "the whistleblower's story - of how two brands of the three-in-one measles, mumps and rubella vaccine (MMR) were marketed in the UK (and worldwide) after having been withdrawn in Canada-was well known at the time Wakefield spoke (*Internet video, Wakefield, The Whistleblower, 25 August 2014*).

(20) a. Jeni Barnett: Do you not think, though, that as a parent, I am allowed to make a decision about what I put in my kid's body?

b. Yasmin: Yes. And do you not think that a parent whose child has cancer and is having chemotherapy and has a much lower resistance to measles, mumps and rubella **and things like that**, has a right for their child to go to normal Primary School? (*Jeni Barnett LBC Radio Show, 7 January 2009*).

In the interactional instance in (b), Yasmin advocates the importance of herd immunity[9] for sick children by resorting to the 'there is more' principle, whereby the existence of other similar viral diseases can be easily inferred by most lay listeners. As a token of intersubjectivity, the adjunctive extender thus directly serves as a cue for radio listeners to infer further instantiations of the same disease category. More importantly, it indirectly contributes to expanding the biased argumentation in (a) so as to reveal intentionally hidden knowledge about the benefits of vaccination. As noted by Overstreet (1999, p. 74), this draws attention to the interlocutors' "social difference", rather than "affirming the participants' solidarity", thus highlighting the "dialectical *relations between* discourse and other [social] elements" beyond its internal relations (Fairclough 2013, p. 4; original emphasis).

5. Discussion and conclusion

Research findings have shown that VL use "needs to be considered with reference to contexts and situations when it will be appropriate or inappropriate" (Channell 1994, p.97). In order to account for the specific case of the autism-MMR vaccine controversy, a very wide range of vague categories has been introduced in the study to examine their functional roles in the diachronic process of (re)constructing knowledge across scientific and lay communities. In detail, the scientific subcorpus significantly revealed:

- a limited use of *numerical approximators*, exception made for Deer as a journalist, thus pointing to the conventional scientific practice of providing exact numbers; a more significant use of *vague neutral quantifiers* was found, whereby Wakefield's main purpose was to express tentativeness in order to convincingly attenuate criticisms from the scientific community, and Deer's vague neutrality was employed to objectively highlight arguable MMR-autism correlations;

[9] Herd immunity is a form of immunity resulting from the vaccination of a significant portion of a population (or herd), which provides a measure of protection for those with a weak immune system.

- *epistemic stance markers* were used by Wakefield to signal scientific doubt in medical texts as expected, whereas more subjective stances were taken in the science-by-press conference genre for the persuasive purpose of convincing the audience of his concern about patient safety; *subjective stance markers* were used more by Deer to unmask Wakefield's research procedures and thus question readers' passive acceptance of the issue;
- *placeholders* were used by Wakefield to establish mutual agreement with his public and gain consensus on his important breakthrough, while Deer's intent was to breach the reciprocity of perspectives and arouse readers' suspicion of Wakefield's misconduct.

On the other hand, the popular subcorpus showed that:

- *numerical approximators* were mostly used in the media genres to subtly impress the public audience, but were also used by medical professionals to communicate scientific uncertainty across the lay community; the more recurring use of vague *multal quantifiers* was understood to deliberately create a major openness of meaning, which allowed for alternative justifications of the fraudulent knowledge claim;
- *epistemic stance markers* were found to mainly function as 'boosting' (Hyland 1998b) devices of misleading propositional content, and to mark the epistemic right of different celebrities to claim authority of knowledge. Regardless of their lack of expertise, these figures took advantage of their popularity to actively participate in the irresponsible process of creating new persuasive interpretations of the same fake knowledge, thus showing how they were able to uncontrollably shake "public confidence in the scientific system" (Furman *et al.* 2012, p. 278); consistently, *subjective stance markers* functioned mainly as markers of imprecision to construct pseudoscientific knowledge claims with the dramatic effect of spreading fear about the uncertain risks of the MMR vaccine in the public arena so as to enhance anti-vaccination decisions;
- *general extenders* were mainly of the *disjunctive* type, and thus used for the functional purpose of downgrading information, especially about people who were not found to entirely agree with the MMR-autism claim.

Traditionally, vagueness in medical discourse "demonstrates a scholarly orderliness in th[e] representation of knowledge" (Prince *et al.* 1982, p. 96), and represents an inevitable resource used to reflect degrees of scientific uncertainty (Salager-Meyer 1994), and to protect medical writers' reputation (Hyland 1998a). This study, instead, has highlighted how a manipulative use can be made of this functionality for the deviating scope of concealing unreliable and unethical knowledge claims. Furthermore, the investigation has shown that once these claims cross the boundaries of the public domain, again VL loses its appropriateness, whereby it serves the important purpose of adjusting complex medical knowledge to "a less scientific discourse community" (Varttala 1999, p. 192). In the present case, it was rather used as an instrumental device to propagate mystified knowledge as reliable scientific advancements, thus impacting on the major reduction in vaccination uptake, and contributing to the current concern for the increasing spread of viral diseases.

Bionotes Anna Franca Plastina is Associate Professor of English Language and Translation. Her research focuses on Discourse Analysis, LSP, Applied Linguistics, Psycholinguistics. Her most recent publications include the chapters *Reframing as a Persuasive Method in Public Speech: Beyond Globalized Biodiversity* (2018, John Benjamins) and *Creating Specialised Corpora in English for Medical Purposes: A Study on Learners' Direct Use* (2017, Peter Lang); the articles *Hashtagging for Health Promotion: Constructing Meaning as an #AntibioticGuardian* (2018, *Altre Modernità*) and *Online Health Promotion: The Cross-Cultural Construction of Biopedagogical Discourses of Obesity* (2017, *ESP Across Cultures*).
Rosita Maglie is a Researcher in English Language and Translation at the University of Bari Aldo Moro. Her main research interests are in ESP, Corpus-based Translation Studies, Computer-mediated Communication, Discourse Analysis and Applied Linguistics. Her PhD dissertation, along with other publications, focus on the use of domain-specific corpora specially collected in order to investigate English and Italian Medical Discourse (2004, 2005, 2007, 2009, 2011, 2015, 2017). Her publication include two volumes: *Understanding the Language of Medicine* (2009, Aracne) and *The new Discourse of Healthcare: A Corpus and Discourse Analysis Approach to a Q&A website* (2015, Aracne).

Authors' addresses: annafranca.plastina@unical.it; rosita.maglie@uniba.it

References

Aijmer K. 1997, *I think - an English modal particle*, in Swan T. and Jansen Westvik O. (eds.), *Modality in the Germanic Languages*, Mouton de Gruyter, Berlin, pp. 1-47.

Baker P., Gabrielatos C., Khosravinik M., Krzyzanowski M., McEnery T. and Wodak R. 2008, *A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press*, in "Discourse and Society" 19 [3], pp. 273-306.

Bauer M. 1998. *'La longue durée' of popular science, 1830- present*, in Devèze-Berthet D. (ed.), *La promotion de la culture scientifique et technique: ses acteurs et leurs logiques*, Université Paris 7, Denis Diderot, Paris, pp. 75-92.

Calsamiglia H. and Ferrero C.L. 2003, *Role and position of scientific voices: reported speech in the media*, in "Discourse Studies" 5 [2], pp. 147-173.

Channell J. 1994, *Vague Language*, Oxford University Press, Oxford.

Cheng W. and Warren M. 2003, *Indirectness, inexplicitness and vagueness made clearer*, in "Pragmatics" 13 [3], pp. 381-400.

Dubois B. 1987, *Something on the order of around forty to forty-four: imprecise numerical expressions in biomedical slide talks*, in "Language in Society" 16 [4], pp. 527-541.

Égré P. and Klinedinst N. 2011, *Introduction: vagueness and language use*, in Égré P. and Klinedinst N. (eds.), *Vagueness and Language Use*, Palgrave Macmillan, New York, pp. 1-24.

Fairclough N. 2013, *Critical Discourse Analysis: the critical study of language* (2nd ed.), Routledge, London/New York.

Furman J.L., Jensen K. and Murray F. 2012, *Governing knowledge in the scientific community: Exploring the role of retractions in biomedicine*, in "Research Policy" 41, pp. 276-290.

Halliday M.A.K. and Matthiessen C.M.I.M. 2004, *An Introduction to Functional Grammar*, 3rd Ed., Hodder Arnold, London.

Hyland K. 1996, *Writing without conviction? Hedging in scientific research articles*, in "Applied Linguistics" 17 [4], pp. 433-454.

Hyland K. 1998a, *Hedging in Scientific Research Articles*, John Benjamins, Amsterdam.

Hyland K. 1998b, *Boosting, hedging, and the negotiation of academic knowledge*, in "Text-Interdisciplinary Journal for the Study of Discourse" 18 [3], pp. 360-372.

Jerome F. 1989, *Science by press conference*, in "Technology Review" 92, pp. 72-73.

Kolodziejski L. 2014, *Harms of Hedging in Scientific Discourse: Andrew Wakefield and the Origins of the Autism Vaccine Controversy*, in "Technical Communication Quarterly" 23 [3], pp. 165-183.

Mautner G. 2009, *Corpora and critical discourse analysis*, in Baker P. (ed.), *Contemporary Corpus Linguistics*, Continuum, London, pp. 32-46.

Myers G. 1985, *Texts as knowledge claims: the social construction of two biology articles*, in "Social Studies of Science" 15 [4], pp. 593-630.

Myers G. 1989, *The pragmatics of politeness in scientific articles*, in "Applied Linguistics" 10 [1], pp. 1-35.

O'Keeffe A. 2003, *'Like the wise virgins and all that jazz' – using a corpus to examine vague language and shared knowledge*, in Connor U. and Upton T.A. (eds.), *Applied Corpus Linguistics: A Multidimensional Perspective*, Rodopi, Amsterdam, pp. 1-20.

Overstreet M. 1999, *Whales, Candlelight and Stuff Like That, General Extenders in English Discourse*, Oxford University Press, Oxford.
Overstreet M. 2005, *And stuff und so: investigating pragmatic expressions in English and German*, in "Journal of Pragmatics" 37 [11], pp. 1845-1864.
Overstreet M. and Yule G. 1997, *On being explicit and stuff in contemporary American English*, in "Journal of English Linguistics" 25 [3], pp. 250-258.
Overstreet M. and Yule G. 2002, *The metapragmatics of and everything*, in "Journal of Pragmatics" 34 [6], pp. 785-794.
Palmer F.R. 2001, *Mood and Modality* (2nd ed.), Cambridge University Press, Cambridge.
Partington A. 2006, *Metaphors, motifs and similes across discourse types: Corpus-assisted discourse studies (CADS) at work*, in Stefanowitsch A. and Gries S. (eds.), *Corpus-based Approaches to Metaphor and Metonymy*, Mouton de Gruyter, Berlin, pp. 267-304.
Plastina A.F. and Del Vecchio F. 2014, *Diagnostic News Delivery: A Microanalysis of the Use of Shields*, in Zuczkowski A., Bongelli R., Riccioni I. and Canestrari C. (eds), *Communicating Certainty and Uncertainty in Medical, Supportive and Scientific Contexts*, John Benjamins, Amsterdam, pp. 183-200.
Prince E.F., Frader J. and Bosk C. 1982, *On Hedging in Physician-Physician Discourse*, in di Pietro R. (ed.), *Linguistics and The Professions. Proceedings of the Second Annual Delaware Symposium on Language Studies*, Ablex Publishing Corporation, Norwood, NJ, pp. 83-97.
Sabet P. and Zhang G.Q. 2015, *Communicating through Vague Language: A Comparative Study of L1 and L2 Speakers*, Palgrave MacMillan, New York.
Salager-Meyer F. 1994, *Hedges and textual communicative function in medical English written discourse*, in "English for Specific Purposes" 13 [2], pp. 149-171.
Stivers T., Mondada L. and Steensig J. (eds.), 2011. *The Morality of Knowledge in Conversation*, Cambridge University Press, Cambridge.
Varttala T., 1999. *Remarks on the Communicative Functions of Hedging in Popular Scientific and Specialist Research Articles on Medicine,* in "English for Specific Purposes" 18 [2], pp. 177-200.
Vreeman R.C. and Carroll A.E. 2007, *Medical myths*, in "British Medical Journal" 335 [7633], pp. 1288-1289.
Wierzbicka A. 1996, *Semantics: Primes and Universals*, Oxford University Press, Oxford.
Zhang G.Q. 2011, *Elasticity of vague language*, in "Intercultural Pragmatics" 8 [4], pp. 571-99.
Zhang G.Q. 2015, *Elastic Language: How and Why We Stretch our Words*, Cambridge University Press, Cambridge.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 121-145
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p121

POPE FRANCIS'S *LAUDATO SI'* A corpus study of environmental and religious discourse[1]

ERIK CASTELLO, SARA GESUATO
UNIVERSITÀ DEGLI STUDI DI PADOVA

Abstract – This paper explores aspects of the lexico-grammar of religiously oriented environmental discourse produced by a leading religious authority, Pope Francis. It examines the most frequent keywords and keyword clusters of the encyclical letter *Laudato Si'* against popularised updates on scientific and technological advances available on the NASA website. The findings show that *Laudato Si'* draws attention both to how people's behaviour affects the environment and to its relevance to the current political and economic situation. The Letter also calls for a much-needed caring attitude towards the environment, and thus appears to be characterized by the directive communicative function throughout, while presenting a more specific religious slant only in select chapters. The analysis carried out highlights both the topics and the rhetorical goals of the discourse of Pope Francis.

Keywords: environmental discourse; religious discourse; keyword analysis; lexical patterns; Pope Francis's *Laudato Si'*

1. Introduction

The environment is the context where life can emerge, evolve and come to an end. All the elements that are part of, and exist in, the environment, are interconnected by contact and/or chains of action/reaction. Therefore, the environment concerns all living entities.[2] Similarly, how the environment is talked about is relevant to all of us: the ways in which it is (re)presented and discussed may have far-reaching social consequences and political repercussions, both local and global, affecting the living conditions of all

[1] Both authors are responsible for the design of the study. The first author carried out the corpus analysis and wrote (sub-)Sections 3.2, 4.1 and 4.2, while the second author wrote(sub-)Sections 1, 2, 3.1 and 5.

[2] However, the fact that it is named *environment*, roughly meaning 'what surrounds us', suggests that is it not considered as important as humans (Goatly 2000, p. 278).

organisms. This is because “language does not passively reflect reality; language actively creates reality” (Halliday 2001 [1990], p. 179), and the way in which reality is constructed in texts affects our choice of action on that very reality (Goatly 2001 [1996], p. 212).

Environmental discourse is the object of study of ecolinguistics, which explores how natural phenomena are integrated into languages and cultures (Steffensen, Fill 2014), how language and the environment are co-constructed and inter-connected (Döring, Zunino 2014), and how language affects ecologically significant actions (Halliday 2007). Ecolinguistics is a socially committed discipline. It strives to critique ecologically destructive behaviour and to encourage relationships of respect and care for the natural world (Stibbe 2014). Therefore, it studies discourses about ecological matters, but also genres like “industry handbooks, lifestyle magazines and economic textbooks” (Stibbe 2014, p. 108), which can reveal what people consider valuable about the environment, and how they treat one another, other living organisms and the physical context in which they all live (Stibbe 2014, p. 109).

Because of its social commitment, ecolinguistics often examines the discourse of social actors with vested interests in phenomena and practices that impact on the environment, for example how British Petroleum comments on the 2010 oil spill (Alexander 2013), how European colonisers describe the “Indies” discovered by Columbus (Döring, Zunino 2014), and how industries and environmental organizations discuss ozone protection (Gerbig 1997). The discourse produced by such interest groups is worth examining because it shows how stakeholders conceptualise the environment and because it can influence public policy makers.

All discourses about ecological issues and those which have some impact on the environment are worthwhile objects of analysis for ecolinguistics. However, those produced by public figures – e.g. scientists, governmental authorities, religious leaders, and media professionals – deserve special attention. These may affect the conscience and conduct of millions of people: they inform the general public about what is known and understood about the environment; they convey beliefs, values and attitudes relevant to it; and they shape collective behaviour.

Yet, not enough is known about such public environmental discourses. For example, we do not know in much detail yet what information and viewpoints they convey, how their viewpoints are influenced by their authors’ linguo-cultural backgrounds, or to what extent their communicative goals are informed by the roles their authors play in the professional-social world – scientists as investigators and educators, politicians as problem solvers and policy makers, religious authorities as spiritual guides, and journalists as reporters and opinion makers.

In this paper, we want to partly fill the above-mentioned knowledge gap by exploring the ecological discourse of Pope Francis. This is an influential figure on the international political and cultural arena – what he says and does is echoed across the entire globe and becomes relevant to millions of people. He is the leader of the Catholic Church, and thus a theological authority and the supreme missionary of the Christian faith. But as the spiritual guide of millions of believers, Pope Francis is also an educator of and reference point for people's collective conscience. He is known for his opposition to consumerism and his commitment to solidarity. With the promulgation of his Encyclical Letter *Laudato Si'* ('Praise Be To You'; 2015), he has shown his support for safeguarding the well-being of humans and his concern for climate change.

As the following excerpt illustrates, *Laudato Si*' directly appeals to the reader to be pro-active about and protective of the environment:

> Our goal is not to amass information or to satisfy curiosity, but rather to become painfully aware, to dare to turn what is happening to the world into our own personal suffering and thus to discover what each of us can do about it. (Laudato Si', Ch. 1, 17).

As far as we know, no linguistic study has been carried out on *Laudato Si'*. However, insightful comments on this publication have been put forward by novelist and thinker Amitav Ghosh. In his non-fiction book *The Great Derangement* (2016), Ghosh examines the spectre of an impending environmental catastrophe looming on the horizon of modern society by pointing out the limits and incoherence of human agency in the social, economic, political and cultural arena. In particular, in Chapter 8, the author comments on the stylistic-rhetorical make-up of two public texts on ecological issues, namely Pope Francis's *Laudato Si'* and the *Paris Agreement* on climate change signed by the United Nations in 2015.[3] Ghosh observes how the former shares virtually the same sources and covers the same topics as the latter, but also how it is highly divergent in its formulation and stance. That is, he describes *Laudato Si'* as simple, sober and clear, and accessible to all readers. He also presents it as overtly critical of present-day paradigms, especially that of unlimited economic growth, which gives voice to the excluded masses making up most of humankind. Overall, Gosh argues that *Laudato Si'* is open and direct, as it considers ecological issues within the larger domain of social inequalities, and shows awareness of the limits of human freedom and abilities (Ghosh 2016, pp. 154-159).

The considerations made by Ghosh, which are backed up by quotations from the texts he compares and contrasts, suggest that a systematic comparative-contrastive textual examination of *Laudato Si'* vis-à-vis another

[3] https://unfccc.int/sites/default/files/english_paris_agreement.pdf.

piece of environmental discourse can be revealing and offer insights into its encoding and content, and a fuller understanding of its ideological orientation. This is what we set out to do in this paper.

In order to outline the distinctive lexico-phraseological encoding of *Laudato Si'*, we compare it with the environmental discourse of an organization that is as visible on the international arena as Pope Francis is, namely NASA, the National Aeronautics and Space Administration. This is an agency of the US federal government, which is responsible for aeronautics, space exploration and the study of the solar system. One of its mission directorates, NASA Science,[4] has a specific research and educational orientation: its mission is to understand, and share knowledge of, the universe, especially the composition, movement, physical forces and weather conditions of our solar system and planet.

NASA, therefore, regularly posts updates on scientific and technological advances relevant to the environment. These texts focus on the physical mechanisms of large-scale phenomena involving forces and entities in the world, and present scientists' rational perspective on them. These short reports are technical in content, and descriptive-informative in slant, as the following excerpt on climate change illustrates:

> The potential future effects of global climate change include more frequent wildfires, longer periods of drought in some regions and an increase in the number, duration and intensity of tropical storms. (https://climate.nasa.gov/effects; picture caption)

Although such online material may differ from *Laudato Si'* in drawing attention to different specific topics, projecting different attitudes, and being driven by different priorities and concerns, it is about the same general subject matter, and targets a similarly wide non-specialist readership as the Encyclical Letter. This makes it a suitable term of comparison. We therefore examine the content and encoding of Pope Francis's *Laudato Si'* by contrasting it against a selection of NASA's online texts on global climate change through a corpus-driven investigation of its lexis. Our goal is to explore what its lexical make-up can reveal about its focus (topics) and stance (positioning on the topics).

In Section 2, we provide more background for our study; in Section 3, we outline our data collection procedure and method of analysis; in Section 4 we report and discuss the findings of our investigation; and in Section 5 we draw implications from our research.

[4] https://science.nasa.gov/.

2. Literature review

Ecolinguistics brings together critical discourse analysis and social activism (e.g. Stibbe 2014, p. 119). This discipline investigates how humans relate to one another and to the larger ecological systems (Stibbe 2014), and how linguistic resources and practices may convey destructive vs supportive views of the environment (Alexander 2000; Halliday 2001 [1990]; Heuberger 2007; Mühlhäusler 2003; Stibbe 2014). It aims to reveal how communicative practices reflect, reproduce and sustain given ways of thinking, believing, feeling and acting. Another of its goals is to empower members of society to act upon these cognitive, emotional and cultural practices so as to reduce inequalities. Because it studies "the impact of language on the life-sustaining relationships among humans, other organisms and the physical environment" (Alexander, Stibbe 2014, p. 105), this research domain questions and exposes "the stories that underpin our current unsustainable civilisation", and is "normatively oriented towards preserving relationships which sustain life" (Alexander, Stibbe 2014, p. 105).[5] Therefore, ecolinguistics is not only descriptive-informative – reporting on the communicative practices of the actors that play a role in, and have their voices heard on, the environment – but also action-oriented – aiming at shaping people's behaviour for the preservation and thriving of the environment.

Ecolinguistic studies investigate recurrent communicative choices in texts about, or somehow having a bearing on, the environment. Recurrent linguistic patterns have the cumulative effect of representing the world in a consistent fashion. Identifying, describing and motivating these patterns shows what people come to perceive as the "normal" way of thinking of and attending to the world (Alexander, Stibbe 2014, p. 105). If these repeated phraseologies and conceptualisations are spelt out, they can be questioned, resisted or replaced with alternative ones (Stibbe 2015, pp. 86, 96).

Early on, Halliday (2001 [1990], pp. 192-193) pointed out how frequent verbal descriptions of phenomena may hide social-political meanings. For example, he showed how the systematic representation of growth as positive, including the growth of consumption of goods, conceals from consciousness the destructive power of consumerism, as growth entails the using up of resources. But he also observed that it would be possible to

[5] Ecolinguistics comprises the study of discourse on ecological matters and also "deeper reflections on the theories of language inspired by the holistic paradigm of ecology" (Bang, Trampe 2014, p. 83). The interaction between the two strands of research (i.e. the role of language in the investigation of environmental issues vs. the relationships between language and its socio-cultural environment) is discussed in Fill (1998). The evolution and perspectives of the discipline are discussed in Steffensen and Fill (2014). A specific overview of language ecology is given in Eliasson (2015).

use the power of grammar to redefine the concept (i.e. by replacing *growth* with *negative shrinkage*) so as to modify our way of thinking of, and conduct towards, consumerism. Halliday also raised our awareness of the grammatical representation, in European languages, of natural resources as existing without limit and of environmental phenomena as inanimate entities. Indeed, he illustrates how, typically, natural resources are encoded in uncountable nouns (e.g. water) and how environmental phenomena are not construed as active participants in events (i.e. as doers).

Most ecolinguistic research examines how interest groups argue for and propagate selected positions with regard to environmental issues, discussing them from different frames (e.g. scientific, political or moral), and giving visibility to or concealing different claims about them. This way, it shows how the environment can be conceptualised as a resource to be exploited or as a site of beauty inhabited by living creatures. Previous works have considered various topics, including: ecological crisis communication produced by multinationals (Alexander 2009, 2013); ecotourism discourse (Stamou, Paraskevopoulos 2008; Mühlhäusler 2000); the representation of nonhuman animals across genres (Gilquin, Jacobs 2006); the representation of anthropocentrism and speciesism (Stibbe 2003); the description of nature on the radio (Goatly 2002); climate change discourses in newspapers (Grundmann, Scott 2014; Grundmann, Krishnamurty 2010) and commercial science journalism (Molek-Kozakowska 2018); the conceptualisation of ‘green’ in news discourse (Bevitori 2011a); environmental issues in the multisemiotic news story genre (Bednarek, Caple 2010); debates on environmental issues in texts produced by industry and environmentalists (Gerbig 1997; Poole 2018); and the role of the environment in American presidential speeches (Bevitori 2011b).[6]

Given its diversified research focus (Chen 2016), ecolinguistics has explored manifold communicative practices in environmental discourse, including: narratives projecting different types of identity; descriptions of situations and reports of events with different degrees of factivity; representation of the degree of involvement of entities in situations/phenomena; the frequency of reference to, naming practices of, and degree of prominence of entities/phenomena in discourse (Stibbe 2015, Ch. 8); lexico-stylistic devices for rendering environmental information newsworthy and appealing (Molek-Kozakowska 2017), and also evaluative, narrative and framing strategies for obscuring it (Molek-Kozakowska 2018). From the lexico-grammatical point of view, ecolinguistic analyses have investigated: collocations, keywords and clusters (Gerbig 1997; Grundmann,

[6] References to additional studies in environmental discourse are found in Grundmann and Scott (2014, p. 221), Poole (2018, pp. 356-357) and Steffensen and Fill (2014, pp. 11-12).

Krishnamurty 2010; Grundmann, Scott 2014; Goatly 2002; Bevitori 2011a, 2011b); pronoun usage (Gilquin, Jacobs 2006); metaphor (Alexander 2013; Döring, Zunino 2014); the representation of agency (or lack thereof) through nominalisations (Gerbig 1997; Goatly 2001 [1996]), nominal compounding (Goatly 2001 [1996]), active/ergative/passive constructions (Gerbig 1997) and transitivity patterns (Goatly 2002); the representation of causation and responsibility through an analysis of ergative/reporting/attitudinal verbs and keywords (Gerbig 1997); explicit and implicit evaluative patterns (Bevitori 2011a; Bednarek, Caple 2010); and distancing techniques (Stibbe 2003).

The methods used include corpus linguistic techniques, critical discourse analysis, content analysis, and the application of the tools of systemic functional linguistics, possibly in combination. In this paper, we use corpus linguistic techniques to explore Pope Francis's view on the environment as conveyed in his *Laudato Si'*.

3. Data and method

Below we describe how we collected the data for our study, and report on how we carried out our analysis.

3.1. Data collection

The data we considered for our analysis comprises the English version of *Laudato Si'* downloadable from the Vatican website[7] and popularised updates on scientific advances available from the NASA website.[8]

Laudato Si' is an 82-page text with the following structure:

- a 5-page untitled introduction;
- Chapter 1: What is happening to our common home
- Chapter 2: The Gospel of creation
- Chapter 3: The human roots of the ecological crisis
- Chapter 4: Integral ecology
- Chapter 5: Lines of approach and action
- Chapter 6: Ecological education and spirituality
- A prayer for our earth
- A Christian prayer in union with creation

[7] http://w2.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html

[8] https://climate.nasa.gov/

- a 10-page list of references.

The introduction is divided into numbered paragraphs, some of which are grouped under headings. The chapters also comprise numbered paragraphs, which are grouped under headings and sub-headings. Altogether, the Encyclical Letter includes 246 paragraphs, the last one ending with two prayers. The References appear in the form of numbered end-notes. For our analysis, we considered the text from the title page to the prayers included.

The home page of the NASA website presents six menus (i.e. *Facts, Articles, Solutions, Explore, Resources* and *NASA Science*). Each is divided into sub-menus (e.g. *Facts* comprises *Evidence, Causes, Effects, Scientific Consensus, Vital Signs* and *Questions (FAQs)*), which present multi-modal updates on given topics. For our research, we considered the content of the *Facts* and *Articles* menus, which provide information about scientists' findings, and of the *Solutions* menu, which offers information on climate change. We chose to do so for two reasons: on one hand, we were interested in *verbal* accounts of environmental issues, which were prominent in the *Facts, Articles* and *Solutions* menus; on the other, we wanted to collect an amount of data comparable in size to that of *Laudato Si'*.

From the first 5 sub-menus of the *Facts* menu we downloaded the text appearing on the left-and-central portion of the page under titled sections. From the last sub-menu we selected the text of each question and the answer accessible after clicking on it, and in the first question, also the text of the second link listed under the heading *More*.

Under the *Articles* menu, we collected 18, 3 and 1 texts, respectively, from the *News*, *Blogs* and *Fun Facts* sub-menus. For each text from the *News* and *Blogs* sub-menus, we considered: the title, the first paragraph sentence, and the rest of the text body accessible by clicking on the title. From the *Fun Facts* sub-menu we selected the material accessible by clicking on the various links listed there.

From the *Solutions* menu, we selected the texts found in the *Mitigation and Adaptation* and *Government and Resources* sub-menus, and the text accessible through the first link in the *Energy Innovations* sub-menu.

We then deleted texts that appeared more than once. In the end, we gathered 6 texts from the *Facts* menu, 22 from the *Articles* menu, and 3 from the *Solutions* menu, for a total of 31 texts.

All the above texts were compiled into two corpora: *Laudato Si'* and the NASA corpus. Table 1 shows the size of the corpora in words and sentences. *Laudato Si*' is made up of 37,730 tokens and 4,878 types, and includes 1,520 sentences, while the NASA corpus comprises 37,840 tokens and 4,744 types, and instantiates 1,753 sentences. The two corpora display highly comparable type/token ratio values (12.92 vs 12.67, respectively), which suggests that they are characterised by similar lexical variation. A difference emerges at the level of syntax, *Laudato Si'* being more elaborate than the NASA corpus, as revealed by their average sentence length (24.82 vs 21.58, respectively).

Statistics	*Laudato Si'*	NASA corpus
Types	4,878	4,744
Tokens	37,730	37,840
T/T ratio	12.92	12.67
Sentences	1,520	1,753
Average words per sentence	24.82	21.58

Table 1
Types, tokens, type/token ratios, number of sentences and average words per sentence in *Laudato Si'* and the NASA corpus.

The two corpora are therefore quite similar in size and lexico-syntactic elaboration. However, given their different focus, that is, the topics addressed and general communicative goals (see Section 1), we thought that the NASA corpus could serve as a useful term of comparison for *Laudato Si'*. In this study we therefore used it as our reference corpus. As observed by Gabrielatos (2018), a reference corpus does not have to be larger than the study corpus or to be a general corpus (see also McEnery 2016; Baker 2006).

3.2. Research questions and software

The analysis of the corpora considered was informed by the following research questions:

What topics are discussed in *Laudato Si'*?

Are the main topics discussed across all the chapters or only in select chapters of *Laudato Si'*?

What lexico-grammatical patterns mainly characterise the discourse of *Laudato Si'*?

We operationalised the above questions as follows:

What keywords characterise *Laudato Si'*, and what semantic fields do they belong to?

In which chapters of *Laudato Si'* do these keywords show up, and with what frequency?

What phraseological patterns of modality are revealed in *Laudato Si'* by clusters of select keywords?

The concordancing software chosen to explore the corpora is the free program *AntConc*, v. 3.5.7 (Anthony 2018). It provides a variety of tools, of which those mainly used to conduct this research were Keywords, Clusters, Concordancer and Concordance Plots. In order to generate the keyword lists for analysis, we opted for the recommended keyness measure “Log-likelihood (4-term)” and keyword statistic threshold $p < 0.05$ (+Bonferroni).[9] After consulting recent research on “Effect Size metrics” (e.g. Gabrielatos 2018) and experimenting with the “Gabrielatos and Marchis%DIFF” metric embedded in the software, we decided to list the keywords we obtained in order of keyness by log-likelihood, because either parameter returned virtually the same list of keywords, albeit partly in a different order.

4. Findings and discussion

The exploration of *Laudato Si’* starts with the study and qualitative interpretation of the keywords obtained from its comparison to the NASA corpus and an overview of the topics dealt with in it (Section 4.1). It then compares the internal sub-division of *Laudato Si’* through the examination of keyword lists and concordance plots (Section 4.2). Finally, it delves into the phraseology of modal expressions through the study of clusters of keywords (Section 4.3).

4.1. The main topics

The technical procedure outlined in Section 3.2 generated 140 keywords. We checked the keyword list independently and assigned each keyword to one main semantic field among the following twelve: *Author*, *Ecology, Economics, Humanity, Mental processes, Modality, Society, Spatial context, Supernatural*, *Textual relations, Time frame* and *Values.* To make our decisions, we examined the concordance lines of each keyword occurrence. We disregarded thirteen keywords that we could not confidently assign to any of the above semantic fields and/or which we did not reach consensus on, and thus considered 127 keywords.

Table 2 shows the keywords subdivided according to the semantic fields they were associated with. The first column displays the fields in order of decreasing textual prominence, depending on the number of keywords associated with them, while the third column presents the keywords relevant to each field in decreasing order of keyness.

[9] This *p* value is suggested by Baker (2006, p. 126) among others.

Semantic fields	*No. of keywords*	*Keywords*
Values	20	love, life, care, good, responsibility, meaning, beauty, integral, dialogue, dignity, faith, respect, sense, freedom, peace, spirituality, attention, justice, creativity, heart
Society	17	social, problems, relationship, others, political, society, forms, church, culture, cultural, family, lifestyle, power, lack, politics, education, violence
Textual relations	17	which, of, who, to, be, not, their, all, and, yet, with, themselves, towards, for, order, cf, thus, (reality)
Economics	16	countries, progress, economic, development, poor, consumption, interests, crisis, production, developing, develop, financial, economy, technological, paradigm, growth
Humanity	15	us, human, beings, we, he, our, humanity, man, ourselves, capable, personal, person, your, relationships, everyone, (him), (his)
Ecology	14	creation, nature, environment, creatures, ecological, environmental, common, resources, ecology, created, creature, things, degradation, her
Supernatural	12	god, his, lord, him, saint, praise, jesus, creator, himself, father, divine, spirit, (he)
Modality	6	must, cannot, need, needs, should, can
Mental processes	4	concern, consider, acknowledge, entails
Spatial context	3	world, reality, universe
Time frame	2	present, often, (yet)
Author	1	I

Table 2
The keywords in *Laudato Si'* (Note: keywords listed twice are relevant to two fields. They appear in parentheses under the field they are less frequently associated with).

The keywords identified and their semantic fields reveal the focus of Pope Francis's discourse and style. The semantic field *Values* contains the largest number of keywords. These not only represent fundamental spiritual and religious values for Christians, but are also relevant to all human beings, who are responsible for solving the current ecological crisis. When addressing societal issues (see category *Society*), the Pope considers such phenomena as politics, culture, education and the Church, and points out the various problems society is affected by, the most prominent of which appears to be violence. Of course, economics is also part of society, but in the Letter it appears to deserve an independent status as a semantic field: specific reference is made to countries that should strive for progress, development, economic and spiritual growth, and should take action in favour of the poor. The field of *Humanity*, of course, encompasses all of us human beings, and that of *Ecology* shows the Pope's concern for the natural world. Together, they indicate how everyone should be concerned with ecology in all its aspects and fight against environmental degradation. Unsurprisingly, the Pope professes his faith in the *Supernatural*, the next semantic field in the table. He makes reference to God – *God* being the second highest ranking keyword in the corpus overall – to the Holy Spirit and to the teaching of Jesus and of various saints.

A clarification needs to be made about the pronoun forms *he*, *him* and *his* associated with both the *Supernatural* and the *Humanity* semantic fields. The pronoun form *he* is used 88 times altogether, of which 34 times (38%) with reference to either God or Jesus and 54 times (62%) with reference to other entities, mainly man (i.e. all human beings), Saint Francis, previous popes and biblical figures. We, therefore, decided to list *he* under the category *Humanity* and to include it in the *Supernatural* one in brackets. By contrast, the pronoun forms *his* and *him* predominantly refer back to God and Jesus (67% and 90% of the times, respectively), and we therefore assigned them to the category *Supernatural* and put them in brackets under *Humanity*.

The remaining semantic fields shed light on other aspects of the Encyclical Letter. For example, the keywords in *Textual relations* include the pronoun *which*, which scores the highest keyness value and ranks among the most recurrent words in the text. A look at a selection of concordance lines in Figure 1 shows that this pronoun is used in both non-restrictive and restrictive relative clauses.

DNA, and many other abilities	which	we have acquired, have given
basis of the Genesis account	which	grants man "dominion" over the
can encourage ways of acting	which	directly and significantly affect
proposals for dialogue and action	which	would involve each of us
be sufficient to prevent actions	which	affect the environment because, wh
limits of this human activity,	which	is a form of power
extreme consumerism and affluence	which	makes it difficult to develop
sually produce specious agreements	which	fail to inform adequately and
spite the international agreements	which	prohibit chemical, bacteriological
ls to internationalize the Amazon,	which	only serve the economic interests

Figure 1
Concordance lines for *which*.

The recurrent use of *which* in restrictive relative clauses – where the pronoun *that* or the zero pronoun could have been used to refer to the object of the clause – represents a mark of stylistic formality in the papal document. Also the keyword *yet* mainly performs a textual function and thus contributes to the *Textual relations* semantic field. It is used as a conjunction of clauses or as a coordinator within phrases in 39 out of 45 cases (87%) and only 6 times (13%) as a time adverbial. Figure 2 provides a selection of concordance lines illustrating these three uses.

Like *yet*, also the conjunctive adjunct *thus* and the abbreviation *cf* contribute to the formality of the Pope's writing style. Finally, the keyword *reality* is used 41 times in the document. More specifically, it is used four times (9%) as part of the linking adverbial *in reality*, while in 37 other instances (91%) it expresses the concept of what actually happens or is true.

has humanity had such power over itself,	yet	nothing ensures that it will be used
subject to the laws of the market.	Yet	access to safe drinkable water is a
will have to be compensated for by	yet	other techniques which may well prove harmful.
more common. Laws may be well framed	yet	remain a dead letter. Can we hope,
new start, but we have not as	yet	developed a universal awareness needed to achieve
consume and destroy, while others are not	yet	able to live in a way worthy

Figure 2
Concordance lines for *yet.*

In the semantic field *Time frame*, the keywords *present* and *often* indicate that this document is oriented towards the description and critical evaluation of the present time frame and of habitual rather than future events. Figure 3 illustrates this tendency and shows that *present* is mainly used as a predicative or attributive adjective:

of the Catholic Church, 310. God is intimately	present	to each being, without impinging on the
nowhere else to go. The impact of	present	imbalances is also seen in the premature
analysis of realities in the service of	present	needs, is at work whether resources are
to see the deepest roots of our	present	failures, which have to do with the
inability to broaden the scope of our	present	interests and to give consideration to those
a licence to carrying on with our	present	lifestyles and models of production and
have attempted to take stock of our	present	situation, pointing to the cracks in the
crisis; to avoid any deterioration of the	present	crisis and the greater imbalances that would
it allow for the radical change which	present	circumstances require. Rather, it may simply

Figure 3
Concordance lines for *present.*

In fact, the Letter also provides indications about the future, and does so through the extensive use of the expressions of obligation/necessity *must*, *cannot*, *need*, *should*, *can,* which make up the field *Modality.* Their presence indicates that the document is characterised by a directive function. Through these items the Pope encourages his readers to take action to solve environmental problems and grow spiritually. The concordance lines or *need* in Figure 4 give a flavour of this call for action (for more details on modality, see Section 4.3):

in these areas. 151. There is also a	need	to protect those common areas, visual landmarks
economies. In this context, there is a	need	for common and differentiated responsibilities.
urban landscape. Many specialists agree on the	need	to give priority to public transportation. Yet
of progress and human development. But we	need	to grow in the conviction that a
transcending immediate economic interest. We	need	to stop thinking in terms of "interventions"
free of the obsession with consumption. We	need	to take up an ancient lesson, found

Figure 4
Concordance lines for *need.*

The semantic field we labelled *Author* is only made up of the frequently employed first person pronoun *I*, which stands for the author of the Letter, viz. the Pope. Furthermore, the processes of cognition listed under *Mental processes* (i.e. *concern, consider, acknowledge, entails*) indicate that the author invites his readers to become aware of our current environmental and spiritual situation and to ponder over it, in the attempt to find ways to improve it. Finally, *Spatial frame* suggests that the Pope's discourse is not limited to *reality* and this *world*, but extends to the *universe.*

4.2. Dispersion of keywords

Judging from the titles of its chapters, *Laudato Si'* appears to have at times a stronger economic and ecological slant (i.e. Chapters 1, 3, 4 and 5), and at other times a more religious and pastoral thrust (i.e. Chapters 2 and 6). In an attempt to ascertain whether this was the case, we first obtained keyword lists for groups of chapters. We compared the word lists for Chapters 1, 3, 4 and 5 (21,591 tokens; 3,590 types) with those for Chapters 2 and 6 (12,914 tokens; 2,513 types) and the other way round. Subsequently, we generated concordance plots to visually represent the distribution of specific keywords.

The left side of Table 3 lists the 8 keywords obtained for Chapters 1, 3, 4 and 5 (vs Chapters 2 and 6), while the right side presents the first 10 keywords (out of 31) for Chapters 2 and 6 (vs Chapters 1, 3, 4 and 5) and their frequencies. The keywords are given in decreasing order of keyness.

Ch. 1, 3, 4 and 5 vs Ch. 2 and 6		***Ch. 2 and 6 vs Ch. 1, 3, 4 and 5***	
Frequency	*Keywords*	*Frequency*	*Keywords*
47	countries	132	god
31	technology	40	you
29	energy	55	his
44	problems	55	he
39	resources	28	him
26	pollution	120	us
23	economy	27	your
21	technological	40	creatures
		26	lord
		40	love

Table 3
The keywords for Chapters 2 and 6 and those for Chapters 1, 3, 4 and 5.

As can be noticed, the subdivision of keywords in Table 3 lends support to the observation that Chapters 1, 3, 4 and 5 concern technological, economic and pollution issues, and that Chapters 2 and 6 are mainly about religious matters.

Concordance plots visualise the position of the various occurrences of a given term in the text. In this study we used them to check whether the keywords *countries*, *technology*, *problems*, *God*, *you*, *his*, *he*, *we*, *human* and *world* are used in the whole text or only in some chapters. As can be seen from the three plots in Figure 5, the words *countries*, *technology* and *problems* are clearly more recurrent at the beginning of the Encyclical Letter (Chapter 1), are rarely used in the following chapter, and are resumed in Chapters 3 to 5. This finding thus lends support to the hypothesis that Chapters 1, 3, 4 and 5 are mainly concerned with economic and ecological issues.

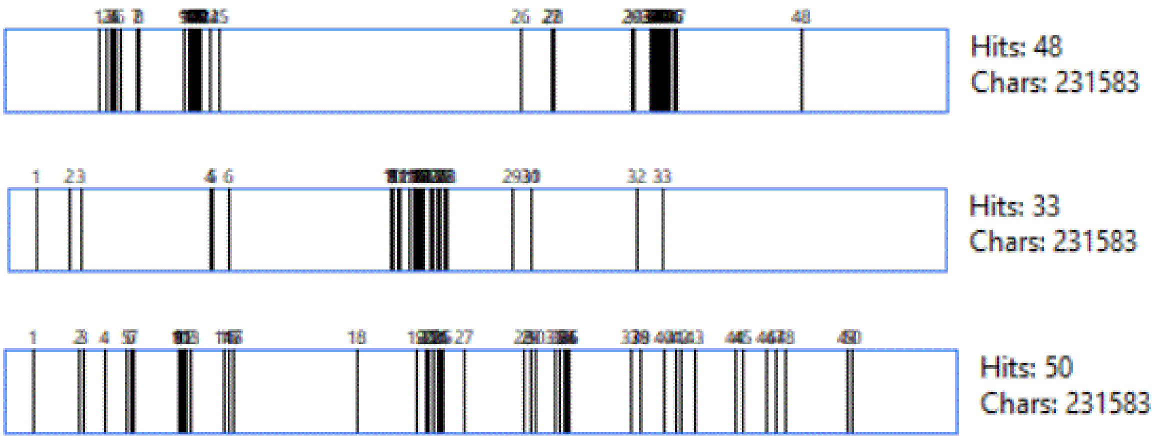


Figure 5
Concordance plots for *countries*, *technology* and *problems* in *Laudato Si'*.

Extract (1) is taken from Chapter 1 and shows how some of the above keywords are used in context:

(1) Some forms of pollution are part of people's daily experience. Exposure to atmospheric pollutants produces a broad spectrum of health hazards, especially for the poor, and causes millions of premature deaths. People take sick, for example, from breathing high levels of smoke from fuels used in cooking or heating. There is also pollution that affects everyone, caused by transport, industrial fumes, substances which contribute to the acidification of soil and water, fertilizers, insecticides, fungicides, herbicides and agrotoxins in general. Technology, which, linked to business interests, is presented as the only way of solving these problems, in fact proves incapable of seeing the mysterious network of relations between things and so sometimes solves one problem only to create others. (our underlining)

The four plots in Figure 6 represent the dispersion of *God*, *you*, *his* and *he*, four keywords denoting *Supernatural* entities. As can be seen by comparing them against those in Figure 5, they occur in portions of the texts other than those where economic and ecological matters are discussed. Specifically, they are very frequently used in Chapters 2 and 6, which indicates that these chapters principally address religious issues.

Figure 6
Concordance plots for *God, you, his, he* in *Laudato Si'*.

Extract (2) from Chapter 2 illustrates the recurrent use of some of these keywords as well as the religious orientation of the chapter:

(2) We are not God. The earth was here before us and it has been given to us. This allows us to respond to the charge that Judaeo-Christian thinking, on the basis of the Genesis account which grants man "dominion" over the earth (cf. Gen 1:28), has encouraged the unbridled exploitation of nature by painting him as domineering and destructive by nature. This is not a correct interpretation of the Bible as understood by the Church. Although it is true that we Christians have at times incorrectly interpreted the Scriptures, nowadays we must forcefully reject the notion that our being created in God's image and given dominion over the earth justifies absolute domination over other creatures. The biblical texts are to be read in their context, with an appropriate hermeneutic, recognizing that they tell us to "till and keep" the garden of the world (cf. Gen 2:15). (our underlining)

The keywords we have discussed so far can be called "local" keywords (Scott, Tribble 2006, p. 66), as they are not used consistently throughout the text. However, in *Laudato Si'* there are also "global" keywords, that is, keywords dispersed throughout the text. This is the case, for example, of the inclusive pronoun *we* and of the words *human* and *world*, as can be seen from Figure 7.

Figure 7
Concordance plots for *we*, *human* and *world* in *Laudato Si'*.

To conclude, the lists of keywords obtained from the comparisons of different groups of chapters in *Laudato Si'* and the concordance plots that visualise their dispersion suggest that the text as a whole calls all human beings to action. However, Chapters 1, 3, 4 and 5 delve specifically into various aspects of current environmental issues, while Chapters 2 and 6 explore the religious implications of the environmental crisis.

4.3. Textual patterns of modality

In this section we explore modality in *Laudato Si'*. We first explore clusters containing expressions of modality in combination with the general keyword *we*, one of the most frequent words in the document and a global keyword. We then focus on modal expressions with grammatical subjects other than *we*, which we thought would be particularly revealing of the author's stance. To this end, we used the Clusters Tool in *AntConc*, which allows one to identify strings of words that recurrently cluster together through a bottom-up approach and makes it possible to gain information about the immediate co-text in which terms are used.

Table 4 presents the thirty most recurrent 2-word clusters beginning with *we* with a minimum frequency of 3 occurrences. It shows that this pronoun is often followed by *can, need, cannot, must, should, would, may* or *will.* At other times, it is used as the subject of predicates containing the verbs of cognition/affection *see*, *know*, *learn*, *think*, *want*, *feel*, *consider* and *look.*

Rank	Freq.	Clusters	Rank	Freq.	Clusters	Rank	Freq.	Clusters
1	43	we are	*11*	6	we would	*21*	4	we not
2	34	we can	*12*	5	we fail	*22*	4	we think
3	34	we have	*13*	5	we may	*23*	4	we want
4	24	we need	*14*	5	we were	*24*	3	we believers
5	13	we see	*15*	5	we will	*25*	3	we consider
6	8	we cannot	*16*	4	we all	*26*	3	we feel
7	8	we must	*17*	4	we continue	*27*	3	we lack
8	8	we should	*18*	4	we find	*28*	3	we look
9	7	we do	*19*	4	we learn	*29*	3	we ourselves
10	7	we know	*20*	4	we live	*30*	3	we share

Table 4
The most recurrent 2-word clusters with the pronoun *we.*

Expressions starting with *we* + a modal verb serve to call attention to options and constraints in collective human actions and experiences, and especially collective responsibility towards the environment. *We can* expresses Pope Francis's conviction that human beings have the ability to do something (e.g. *we can redirect our steps*) or no longer have the option to do something (e.g. *we can no longer view reality in a purely utilitarian way*), while *we cannot* indicates that we are unable to do something (e.g. *we cannot control them)*. With the cluster *we need/must/should* he attempts to direct us towards a given behaviour (e.g. *we need also to think of containing growth*; *we must regain the conviction that we need one another*; *we should be concerned for future generations*), while with *we may* he predicts the likelihood that specific events, especially negative ones, will take place (e.g. *we may well be leaving to coming generations debris, desolation and filth*).

Table 5 displays the first twenty 3-4 word clusters around *we*. Most of them contain the same modal expressions and verbs of cognition/affection which we have just discussed. However, another modal expression emerges from this list: *we are called to.*

Rank	Freq.	Clusters	Rank	Freq. (cont.)	Clusters
1	14	we need to	*11*	3	we can no
2	6	we are called	*12*	3	we can no longer
3	5	we are called to	*13*	3	we can see
4	5	we do not	*14*	3	we continue to
5	5	we fail to	*15*	3	we have to
6	4	we know that	*16*	3	we need only
7	4	we learn to	*17*	3	we see this
8	4	we see how	*18*	3	we should be concerned
9	4	we should be	*19*	3	we should not
10	4	we want to	*20*	3	we would be

Table 5
The most recurrent clusters with the pronoun *we*.

We are called to is a near-modal expression of obligation, which represents yet another linguistic realisation of the Pope's call for commitment to ecology and ecological spirituality. Figure 8 shows the concordance lines for *we are called to,* illustrating how people should take action. That is, *we* all should: acknowledge our *contribution* (to the disfigurement and destruction of creation); be *instruments of God*; *recognize that other living beings* (have a value of their own in God's eyes); *respect creation*; *include in our work a dimension of receptivity* (and gratuity).

generate small ecological damage",	we are called to	acknowledge "our contribution
the last two hundred years. Yet	we are called to	be instruments of God
to use the earth's goods responsibly,	we are called to	recognize that other living beings
dignity and our gift of intelligence,	we are called to	respect creation and its inherent laws
important about work: its meaning.	We are called to	include in our work a dimension of receptivity

Figure 8
Concordance lines for *we are called to.*

Besides *are called to*, the clusters in Table 5 show additional near-modal expressions. These include *we do not*, *we fail to*, which indicate inability to experience/act or conscious choice not to act, and also *we learn to*, *we see how,* and *we see this*, which indicate ability to experience phenomena.

After examining the patternings of *we* with the above modal and near-modal verbs, we also checked whether these modal expressions occur with other subjects. To do so, we searched the Encyclical Letter for 2-word clusters obtained from *AntConc* by setting the search term position to the right. Table 6 shows the most recurrent clusters for the key modal expressions *must, can(not), need(s)* and *should* with a minimum frequency of 1 occurrrence.

	Must		*Can(not)*		*Need(s)*		*Should*	
Rank	Fr.	Clusters	Fr.	Clusters	Fr.	Clusters	Fr.	clusters
1	8	we must	42	we can/cannot	24	we need	8	we should
2	6	it must	19	which can/cannot	5	they need	4	it should
3	2	he must	6	it can/cannot	3	it needs	3	this should
4	2	politics must	3	one can	1	approach needs	2	which should
5	2	society must	3	problems cannot	1	attention needs	1	assessment should
6	2	which must	3	they can	1	debate needs	1	believers should
7	1	account must	3	who can	1	efforts need	1	consensus should
8	1	attention must	2	community can	1	places need	1	creativity should
9	1	beings must	2	education can	1	politics needs	1	education should
10	1	consideration must	2	environment can	1	questions need	1	goal should

Table 6
The most recurrent 2-word clusters with modal verbs on the right position.

As can be seen, some other pronouns other than *we* tend to rank high on the lists (e.g. *it, he, they, which, one, this*). The modal verbs, however, are also used in combination with subject noun phrases denoting human institutions or activities (e.g. *politics, society, community, education*), groups of people (e.g.

beings, *believers*), or cognitive processes or procedures (e.g. *attention*, *consideration*, *approach*, *assessment*, *efforts*).

Overall, the modal expressions identified and their patterns of occurrence suggest that *Laudato Si'* is mainly oriented towards the expression of deontic modality, qualifying the degree of human involvement in and responsibility for the well-being of the planet. Additionally, the findings show that the text draws attention to the possibility for humankind to perceive and become aware of the planet's present condition and future prospects.

5. Conclusion

In this study, our goal was to describe the lexico-phraseological patterns in the communicative practices of an environmentally conscious, religion-inspired publication targeting a lay audience, with a view to revealing how it attends to ecological matters. We set out to outline recurrent, salient topics and notions in *Laudato Si'* and how these are represented by comparing it against a corpus of NASA texts on similar topics and of similar size.

The keyword analysis shows that *Laudato Si'* covers the domain of the environment (*Ecology*) as intimately interconnected with other global social phenomena (*Society*, *Economics*) and as relevant to humankind (*Humanity*), its physical context (*Spatial context*) and the divine (*Supernatural*). Environmental issues are also presented as important for the immediate present and future circumstances of humanity (*Time frame*), and described as part of a crucial aspect of moral consciousness (*Values*). The general stance of the text is that of a formal, yet reader-friendly (*Textual relations*) and collaborative communicative process, in which the Pope (*Author*) directly involves the readership in an awareness-raising initiative (*Modality*, *Mental processes*) on the value of the environment and people's collective responsibility for it. All these semantic domains appear to be discussed with everyday rather than technical/sophisticated terms (e.g. *lifestyle, love, world*).

Laudato Si' also appears to have a narrative development, in the sense that not all its chapters cover the same topics, judging from its lexis. The dispersion of the keywords across the text signals, on one hand, its thematic uniformity, and on the other, its internal textual boundaries (i.e. topic shifts). *Laudato Si'* appears to stress the bonds of communion of all creatures from beginning to end; yet, while Chapters 1, 3, 4 and 5 are strongly focused on environmental issues (see the prominence of such keywords as *energy*, *pollution* and *resources*), Chapters 2 and 6 are more reflective of the relationship between human beings and the divine, characterized by benevolence (see the prominence of such keywords as *God*, *creatures* and *love*).

Finally, the keyword cluster analysis shows that *Laudato Si'* recurrently refers to relationships, principles and processes collectively involving, affecting or appealing to human beings as members of interrelated social groups. Indeed, *Laudato Si'* highlights the interactive dimension of environmental discourse (see the frequent use of inclusive *we*), calls attention to the need for a caring attitude towards the environment (see the use of modals of necessity/obligation), and focuses on the current and prospective time frame (see the use of the present tense and modal verbs).

Overall, the analysis shows how linguistic features recurrently instantiated in the text come to encode a consistent worldview, that is, a stance conveying coherent values, beliefs and opinions (Stibbe 2014, pp. 122-124). *Laudato Si'* frequently refers to the conceptual domains of religious and human values. It is focused on the importance of the bond between human beings, their common home, and the divine. It reflects on the social actors involved in environmental issues, their attitude and actions; it also considers the (risky) processes affecting the natural world, and assesses the interconnectedness of them all.

The Encyclical Letter instantiates a discourse "which could potentially help protect and preserve the conditions that support life" (Stibbe 2014, p. 119) through the presentation of a socially committed approach to ecology. *Laudato Si'* presents data on the natural world with a reflective and interpersonal orientation. It explains to the lay audience how living beings depend on one another for their survival, and it is informed by "an ethical framework to decide why survival and flourishing matters and whose survival and flourishing matters" (Stibbe 2014, p. 119). For its author, the natural world is a crucial element of a shared life-sustaining system.

Our analysis suggests that a lexico-phraseological description of environmental discourse can fruitfully contribute to uncovering the ideas, beliefs and values encoded in it. The findings presented here reveal not only what the text examined is about, but also its rhetorical goals. Confirmation of the present findings could come from widening the scope of the object of analysis – to include, e.g., metaphors and explanation strategies – and/or from applying complementary analytical tools – such as automatic semantic tagging – and/or from carrying out comparative analyses with different datasets (e.g. *Laudato Si'* against the *Paris Agreement*).

Bionotes: Erik Castello (Ph.D., University of Padua, 2007) is associate professor of English language and linguistics at the University of Padua, Italy. His research interests include (learner) corpus linguistics, the language of tourism, systemic functional linguistics and language testing. He has recently examined aspects of written and spoken learner language, including the use of *it*-extraposition constructions and that of discourse markers, and aspects of tourism discourse, including the use of evaluative language and locative constructions. He has published several articles on these topics. He has recently co-edited a volume on Learner Corpus Research.
Sara Gesuato (Ph.D., University of Padua, 1999, University of California at Berkeley, 2004) is associate professor of English language and linguistics at the University of Padua, Italy. Her research interests include pragmatics, discourse and genre analysis, verbal aspect, and corpus linguistics. In her publications, she has discussed the structure and wording of initiating and reacting speech acts, the phraseology and rhetorical structure of academic genres, and the temporal and aspectual meanings of catenative motion verb constructions. She has recently co-edited two volumes on pragmatic issues in language teaching and learning and is currently investigating pedagogical applications of speech act analysis.

Authors' addresses: erik.castello@unipd.it, sara.gesuato@unipd.it

Acknowledgements: The authors are indebted to two anonymous reviewers for helpful comments on a previous version of the paper.

References

Alexander R.J. 2000, *The framing of ecology: some remarks on the relation between language and economics*, in Kettemann B. and Penz H. (eds.), *ECOstructing language, nature and society. The ecolinguistic project revisited*, Stauffenburg, Tübingen, pp. 173-190.

Alexander R.J. 2009, *Framing Discourse on the Environment. A Critical Discourse Approach*, Routledge, New York/London.

Alexander R.J. 2013, *Shaping and misrepresenting public perceptions of ecological catastrophes: the BP Gulf oil spill*, in "Critical Approaches to Discourse Analysis across Disciplines" 7 [1], pp. 1-18.

Alexander R.J. and Stibbe A. 2014, *From the analysis of ecological discourse to the ecological analysis of discourse*, in "Language Sciences" 41, pp. 104-110.

Anthony L. 2018, *AntConc (Version 3.5.7) [Computer Software]*, Tokyo, Japan, Waseda University. http://www.laurenceanthony.net/software (10.01.2018).

Baker P. 2006, *Using Corpora in Discourse Analysis*, Continuum, London.

Bang J.C. and Trampe W. 2014, *Aspects of an ecological theory of language*, in "Language Sciences" 41 [5], pp. 83-92.

Bednarek M. and Caple H. 2010, *Playing with environmental stories in the news – good or bad practice?*, in "Discourse & Communication" 4 [1], pp. 5-31.

Bevitori C. 2011a, *'Jumping on the green bandwagon': The discursive construction of GREEN across 'old' and 'new' media genres at the intersection between corpora and discourse*, in *Corpus Linguistics 2011 – Discourse and Corpus Linguistics*, University of Birmingham, 20-22 July 2011.

Bevitori C. 2011b, *Discursive construction of the environment in American presidential speeches 1960-2013: a diachronic corpus-assisted study*, in McEnery T. and Baker P. (eds.) *Corpora and discourse studies: integrating discourse and corpora*, Basingstoke, Palgrave MacMillan, pp. 110-133.

Chen S. 2016, *Language and ecology: A content analysis of ecolinguistics as an emerging research field*, in "Ampersand" 3, pp. 108-116.

Döring M. and Zunino F. 2014, *Nature Cultures in Old and New Worlds. Steps towards an ecolinguistic perspective on framing a 'new' continent*, in "Language Sciences" 4, pp. 34-40.

Eliasson S. 2015, *The birth of language ecology: interdisciplinary influences in Einar Haugen's "The ecology of language"*, in "Language Sciences" 50, pp. 78-92.

Fill A. 1998, *Ecolinguistics – State of the art 1998*, in "Arbeiten aus Anglistik und Amerikanistik" 23 [1], pp. 3-16.

Gabrielatos C. 2018, *Keyness analysis: nature, metrics and techniques*, in Taylor C. and Marchi A. (eds.), *Corpus Approaches to Discourse: A critical review*, Oxford, Routledge, pp. 225-258.

Gerbig A. 1997, *Lexical and grammatical variation in a corpus. A computer-assisted study of discourse on the environment. Duisburg Papers on Research in Language and Culture, vol. 33*, Peter Lang, Frankfurt am Main.

Gilquin G. and Jacobs G.M. 2006, *Elephants (who) marry mice are very unusual: the use of the relative pronoun "who" with nonhuman animals*, in "Society & Animals" 14 [1], pp. 79-105.

Goatly A. 2000, *Critical reading and writing*, Routledge, London/New York.

Goatly A. 2001 [1996], *Green grammar and grammatical metaphor, or language and myth of power, or metaphors we die by*, in Fill A. and Mühlhäuser P. (eds.) *The ecolinguistics reader. Language, ecology and the environment*, Continuum, London, pp. 203-225.

Goatly A. 2002, *The representation of nature on the BBC World Service*, in "Text" 22 [1], pp. 1-27.

Ghosh A. 2016, *The great derangement. Climate change and the unthinkable,* The University of Chicago Press, Chicago/London.

Grundmann R. and Krishnamurthy R. 2010, *The discourse of climate change: a corpus-based approach*, in "Critical Approaches to Discourse Analysis across Disciplines" 4 [2], pp. 125-146.

Grundmann R. and Scott M. 2014, *Disputed climate science in the media: Do countries matter?*, in "Public Understanding of Science" 23 [2], pp. 220-235.

Halliday M.A.K. 2001 [1990], *News ways of meaning. The challenge to applied linguistics*, in Fill A. and Mühlhäusler P. (eds.), *The ecolinguistics reader. Language, ecology and environment*, Continuum, London, pp. 175-202.

Halliday M.A.K. 2007, *Applied linguistics as an evolving theme*, in Webster J. (ed.), *Language and Education: Collected Works of M.A.K. Halliday*, Continuum, London, pp. 1-19.

Heuberger R. 2007, *Language and ideology. A brief survey of anthropocentrism and specism in English*, in Fill A. and Penz H. (eds.), *Sustaining language*, LIT-Verlag, Münster, pp. 105-124.

McEnery T. 2016, *Keywords,* in Baker P. and Egbert J. (eds.), *Triangulating Methodological Approaches in Corpus-Linguistic research,* Routledge, New York/London, pp. 20-32.

Molek-Kozakowska K. 2017, *Communicating environmental science beyond academia: Stylistic patterns of newsworthiness in popular science journalism*, in "Discourse & Communication" 11 [1], pp. 69-88.

Molek-Kozakowska K. 2018, *Popularity-driven science journalism and climate change: A critical discourse analysis of the unsaid*, in "Discourse, Context & Media" 21, pp. 73-81.

Mühlhäusler P. 2000, *Bleached language on unbleached paper. The language of ecotourism*, in Ketteman B. and Penz H. (eds.), *ECOnstructing Language, Nature and Society. The Ecolinguistic Project Revisited. Essays in Honour of Alwin Fill*, Stauffenburg Verlag, Tübingen, pp. 241-251.

Mühlhäusler P. 2003, *Language of environment – Environment of language*, Battlebridge, London.

Poole R. 2018, *Ecolinguistics, GIS, and Corpus Linguistics for the Analysis of the Rosemont Copper Mine Debate*, in "Environmental Communication" 12 [4], pp. 525-540.

Scott M. and Tribble C. 2006, *Textual patterns: Key words and corpus analysis in language education*, John Benjamins, Amsterdam/Philadelphia.

Stamou A.G. and Paraskevopoulos S. 2008, *Representing protection action in an ecotourism setting: a critical discourse analysis of visitors' books at a Greek reserve*, in "Critical Discourse Studies" 5 [1], pp. 35-54.

Steffensen S. and Fill A. 2014, *Ecolinguistics: The state of the art and future horizons*, in "Language Sciences" 41, pp. 6-25.

Stibbe A. 2003, *As charming as a pig: The discursive construction of the relationship between pigs and humans*, in "Society & Animals" 11 [4], pp. 375-392.

Stibbe A. 2014, *An ecolinguistic approach to critical discourse studies*, in "Critical Discourse Studies" 11 [1], pp. 117-128.

Stibbe A. 2015, *Ecolinguistics. Language, ecology and the stories we live by,* Routledge, London.

Webliography

Laudato Si', http://w2.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html(13.05.2018).
NASA, https://science.nasa.gov/(13.05.2018).
NASA, https://climate.nasa.gov/ (13.05.2018).
NASA, The consequences of climate change. https://climate.nasa.gov/effects(13.05.2018).
Paris Agreement, https://unfccc.int/sites/default/files/english_paris_agreement.pdf (13.05.2018).

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 147-177
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p147

TRUMP IS ERASING CLIMATE CHANGE... LANGUAGE
A corpus-assisted Critical Discourse Analysis of the US online environmental communications under Obama and Trump

ANTONELLA NAPOLITANO, MARIA CRISTINA AIEZZA[1]
UNIVERSITÀ DEGLI STUDI DEL SANNIO

Abstract – Climate change constitutes one of the major challenges of our time. The United States, in particular, represent one of the main greenhouse gas emitters in the world. Under the Obama administration, the US promoted a plan to reduce carbon pollution and incentivise clean energy. A constant stream of information on the impacts of climate change was disseminated online. By contrast, President Trump's election has been linked to a tendency to discredit scientific knowledge. Pursuing an "America first energy plan", Trump aims at rescinding environmental regulations he considers an impediment to business. The present study analyses a corpus of environmental information about global warming and energy policies published on official governmental websites, comparing the data and reports produced during Obama's office with the most recent communications which reflect the priorities of Trump's administration. Applying Corpus-Assisted Critical Discourse Analysis, the study investigates how scientific knowledge about climate change has been reproduced online to serve different interests and support contrasting ideologies. The paper investigates the selection and prominence attributed to specialised information, the argumentations exploited to justify political choices and the authoritative sources quoted to support positions. Particular relevance is assigned to the discourse accompanying the dismissal of basic climate change tenets and the demolition of environmental programmes operated by the current US Presidency. The study shows how Trump's reshaping of environmental policy priorities involves refashioning online contents, by excising, hiding or limiting the importance of any mentions to climate change. Moreover, the present administration is shifting emphasis towards usage of fossil fuels, based on an anachronistic contrast between stewardship of natural resources and economic development.

Keywords: Critical Discourse Analysis; Corpus-Assisted Discourse Analysis; climate change; political discourse; environmental discourse.

[1] The authors discussed and conceived the article together. In particular, Antonella Napolitano is responsible for sections 1, 2, 4.4, 4.5, Maria Cristina Aiezza for sections 3, 4.1, 4.1.1, 4.2, 4.3, 5.

We must not be indifferent or resigned to the loss of biodiversity and the destruction of ecosystems, often caused by our irresponsible and selfish behaviour.
Because of us, thousands of species will no longer give glory to God by their very existence.
We have no such right.
(Pope Francis, 01/09/2016).

1. USA and climate change

Climate change represents one of the main challenges for humankind in this century. The United Nations Intergovernmental Panel on Climate Change (IPCC; UN 2012) periodically assesses international publications by climate scientists "to provide the world with a clear scientific view on the current state of knowledge in climate change and its potential environmental and socio-economic impacts".[2] In one of its last comprehensive studies, the organisation reached a categorical conclusion: unprecedented climate changes observed since the 1950s – such as atmosphere and ocean warming, sea level rise, permafrost reduction – reveal that global warming is "unequivocal" and influenced by the human activity conducted since the Industrial Revolution (IPCC 2013, pp. 4, 17).

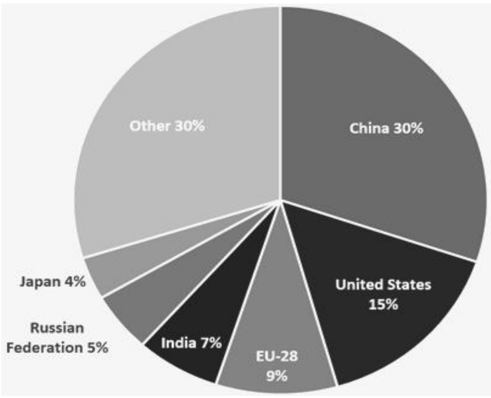


Table 1
Global CO_2 emissions from fossil fuel combustion and industrial processes (Boden *et al*. 2017).

[2] IPCC (United Nations Intergovernmental Panel on Climate Change), *Organization*. https://www.ipcc.ch/organization/organization.shtml (08.03.2018).

Carbon dioxide emissions deriving from fossil fuel combustion and industrial processes represent the largest source of greenhouse gas emissions, leading to an increase in the global temperature. In particular, being one of the most industrialised countries, the United States of America are the second biggest greenhouse gas emitter in the world (see Table 1):

Barack Obama considered climate change as one of the most pressing challenges of our time. The former US President attached particular importance to environmental issues, as testified, for instance, by the words he pronounced in 2013, during his second Inaugural Address. Obama described the destructive effects of climate change in metaphorical terms, as similar to a terrorist invasion:

> We, the people, still believe that our obligations as Americans are not just to ourselves, but to all posterity. We will respond to the threat of climate change, knowing that the failure to do so would betray our children and future generations. Some may still deny the overwhelming judgment of science, but none can avoid the devastating impact of raging fires, and crippling drought, and more powerful storms. [underlining added] (Obama 21.01.2013)

The trope of war here exploited recalls the language typically associated to global warming in the US news discourse, which frequently employs terms such as *threat*, *reduce* or *fight* (Grundmann, Krishnamurty 2010).

During the Obama Presidency, a constant stream of information on the ongoing and projected impacts of climate change was initiated and disseminated online, especially through the website of the Environmental Protection Agency (EPA).[3] A diametrically opposed position is being held by President Donald Trump. His 2017 election was linked to a "growing popular suspicion of expertise", a tendency to consider scientific knowledge as mere opinions and "to seek out alternative narratives to fact-based analysis" (Anthony 18.03.2017). In the past few years, Trump had already been expressing his climate denier views, for instance in over 100 posts on his Twitter social media account (Matthews 01.06.2017). He defined climate change as a 'hoax' and mocked the issue through a quite annoying mix of ignorance and sarcasm, as evident from his *tweets*:

> They changed the name from 'global warming' to 'climate change' after the term global warming just wasn't working (it was too cold)! (Donald J. Trump, @realDonaldTrump 7:15 PM, 25.03.2013)

[3] "Established in 1970 in the wake of elevated concern about environmental pollution, the entity coordinates federal research, monitoring, standard-setting and enforcement activities with the aim of ensuring a cleaner and healthier environment." EPA (United States Environmental Protection Agency), *History*. https://www.epa.gov/history (05.03.2018).

> Ice storm rolls from Texas to Tennessee – I'm in Los Angeles and it's freezing. Global warming is a total, and very expensive, hoax! (Donald J. Trump, @realDonaldTrump 4:13 PM, 06.12.2013)

> It's really cold outside, they are calling it a major freeze, weeks ahead of normal. Man, we could use a big fat dose of global warming! (Donald J. Trump, @realDonaldTrump 2:30 PM, 19.10.2015)

In the US, the terminological distinction between the expressions *climate change* and *global warming* has recently received considerable attention for the perceived value-laden nature of such labels. The lexical issue was even examined by interest groups and political strategists from various parties. The former expression was used to present the environmental problem from a more scientific perspective, while the latter was chosen to depict a more dramatic scenario requiring political action (Grundmann, Krishnamurty 2010, pp. 131-132). With regard to the debate, Frank Luntz, Republican consultant under George W. Bush, in particular, urged that *climate change* be used instead of *global warming*, since

> *"[c]limate change is less frightening than global warming"* […] While global warming has catastrophic communications attached to it, climate change sounds a more controllable and less emotional challenge. (Luntz 2002, p. 142)

A more extreme position is being held by the present Republican administration. In particular, President Trump described environmental laws as an impediment to business and launched his 'America first' energy plan. Since his access to the US Presidency, Trump has been trying to dismantle much of the past legislation on emissions, especially Obama's rules known as *Clean Power Plan*, which limited carbon pollution from US power stations (see also Paragraph 4.4).

President Trump appointed as head of the EPA a climate change sceptic, Oklahoma Attorney General Scott Pruitt, who promised to weaken regulation of carbon emissions from cars and power plants (see also Paragraph 4.3). In the past, Pruitt had actually acted in close concert with oil and gas companies to challenge environmental regulations and, on his LinkedIn profile, he even described himself as a "leading advocate against the EPA's activist agenda" (Milman, Rushe 22.02.2017).

On June 1, 2017, President Trump also announced the United States' withdrawal from the *Paris Climate Agreement*. Signed in 2016 within the United Nations Framework Convention on Climate Change (UNFCCC), the pact aimed to limit global warming and "strengthen the ability of countries to deal with the impacts of climate change".[4] While reaffirming his desire to invigorate domestic coal mining, Trump described the mitigation actions required by the pact as a threat to national interests:

> The Paris Accord would undermine our economy, hamstring our workers, weaken our sovereignty, impose unacceptable legal risks, and put us at a permanent disadvantage to the other countries of the world." (Trump 01.06.2017)

In addition to its indifference towards environmental issues, the administration is also showing a worrying and absurd lack of respect for the work of professionals. As observed by The National Geographic constant observatory, science advisors have been dismissed and online scientific contents have been excised. Hereafter, some titles from its running list of news articles on how Trump is 'changing' the US science:

> "Scramble to save science data"
> "'Science' scrubbed"
> "EPA chief downplays climate"
> "Science and environment budget threatened"
> "EPA scientist retires with a bang"
> "Climate change staffers reassigned"
> "Interior Department scrubs climate change website"
> "Scientists march on Washington"
> "EPA scrubs climate change website"
> "EPA dismisses science advisors" (Greshko *et al*. 25.10.2017)

In particular, by exploiting the volatile feature of the Internet, the present administration is gradually refashioning online contents, by removing, hiding or limiting the importance of any references to climate change in official media, while shifting emphasis away from renewable energy and towards traditional usage of fossil fuels (Milman, Morris 14.05.2017). Worrying reports also come from the Environmental Data and Governance Initiative (EDGI), an international network of academics and non-profits addressing potential threats to federal environmental and energy policy and scientific research. Its Website Monitoring Committee records how data, information, and their presentation is altered in official websites. In particular, the organisation has detected changes in the online pages of: Department of State, Department of Energy; EPA; Department of the Interior; White House; Government Accountability Office; Department of Transportation; Department of Health and Human Services; Federal Emergency Management Agency (see EDGI).

[4] UNFCCC (United Nations Framework Convention on Climate Change), *The Paris Agreement*. http://unfccc.int/paris_agreement/items/9485.php (09.03.2018).

Along the reshaping of environmental policy priorities, Trump is thus also questioning and limiting access to scientific evidence of global warming. Such a reactionary attitude seems to acknowledge the power of language to shape reality, "the constructive effects discourse has upon social identities, social relations and systems of knowledge and belief" (Fairclough 1992, p. 12). Trump's spectacular political turnaround is thus being accompanied by a forced discursive change in the information issued by official government sources. By rubbing out mentions of climate change, the present administration is thus attempting to erase the relevance of the issue in the public opinion.

Considering such shift in the environmental discourse recently experienced in the United States, the present research analyses a corpus of online communications on the issue of climate change comparing texts published during the last Presidencies.

2. Aims and purposes

The study aims at investigating how scientific knowledge about climate change has been reproduced, represented – and altered – online under Obama and Trump Presidencies, focusing on how the US official environmental discourse is being rewritten in order to reflect the priorities of the current Republican administration. More specifically, the paper considers the choice and relevance attributed to specialised knowledge in governmental publications and websites and the argumentations and authoritative sources exploited to justify decisions and policies.

3. Corpus and methods

In order to compare diversified documents produced during the two different administrations, it appeared useful to integrate multiple analytical perspectives, applying a framework (Baker *et al*. 2013) which combines Critical Discourse Analysis (CDA) (see e.g. van Dijk 2011) and Corpus Linguistics (see e.g. Baker 2006). Critical approaches to discourse aim at showing how discourse is both shaped by and contributes to shaping relations of power and ideologies (Fairclough 1992, p. 12). CDA involves the close examination of how language is used in texts to represent particular stances:

> The analysis of representational processes in a text, therefore, comes down to an account of what choices are made – what is included and what is excluded, what is made explicit or left implicit, what is foregrounded and what is backgrounded, what is thematized and what unthematized, what process types and categories are drawn upon to represent events, and so on. (Fairclough

1995, p. 104)

The analysis of the discourse associated with environmental endangerment represents a goal strictly related to the Critical Discourse Studies agenda. Ecological destruction can, indeed, be considered as part of the existing oppressive relations between humans and other humans and between humans and nature (see Stibbe 2014). As Fairclough (2004, p. 104) stated, "[t]he unrestrained emphasis on growth [also] poses major threats to the environment". The capitalist world often downplays the intrinsic ethical value of flora and fauna in the name of the view of nature as a commodity. At the same time, the natural world is also discursively erased from human consciousness, supporting its anthropocentric exploitation and reducing human responsibility in its devastation. Such an 'oblivion' of nature takes place at multiple levels, from sentences and clauses, e.g. through a series of linguistic devices such as metaphors, metonymies, nominalisations, passivisation, ergativity (see e.g. Gerbig 1993; Goatly 2001), to texts and discourses as a whole (Stibbe 2014, pp. 587-588; for a review of the literature on ecolinguistics see e.g. Alexander, Stibbe 2013).

Corpus analysis tools support the present investigation, in that they allow the researcher to identify themes and patterns which may not be evident to the naked eye – e.g. through frequencies, keywords and collocations – thus pinpointing areas for a subsequent close analysis (Baker *et al*. 2013, pp. 20-28). The study also considered how specialised scientific knowledge is popularised, i.e. diffused to the general public (Gotti 2005, p. 203), evaluating how the informative purpose has been bent to promote "private intentions" and ideological interests (Bhatia 2004). Since popularisation involves the transformation of a source text into a derived text, the redrafting can generate an imperfect equivalence. Language and facts may thus be oversimplified, and approximation and omission of specific contents may actually conceal a deliberate distortion.

Through such manifold framework, the study compared a selection of facts, data and reports produced during Obama's office with the most recent communications and modifications by the Trump administration. The corpus is articulated in two main subcorpora, collecting texts issued under the two Presidencies. As shown in Table 2, the corpus is further divided in subsections, organised by source and topic. Both subcorpora include: the documents stating EPA's strategy; the scientific report *National Climate Assessment*; the regulatory measures about the Clean Power Plan published on the EPA websites. The corpus also collects, on the one hand, the climate change information published on the EPA website during the Obama Presidency and, on the other, the news releases on how to comply with Trump's new environmental policies in the sections *Climate* and *Energy* and published between February 2, 2017 and November 9, 2017.

Obama Subcorpus	**Tokens**	**Types**	**Trump Subcorpus**	**Tokens**	**Types**
Total	1,347,252	27,032	Total	323,933	13,997
EPA Strategic Plan 14-18	38,283	3,927	*EPA Strategic Plan 18-22*	14,916	2,099
National Climate Assessment 2014	497,594	19,000	*National Climate Assessment 2017*	272,305	11,266
EPA website: *Clean Power Plan*	771,647	10,921	EPA website: *Complying with President Trump's Executive Order on Energy Independence*	30,514	2,693
EPA website: *Climate Change* Section	59,904	4,564	EPA website: News Releases on *Climate* and *Energy*	12,698	2,126

Table 2
Corpus overview.

As evident from the data displayed in Table 2, the two subcorpora may not appear to be exactly comparable. This is mainly due to the limited amount of materials recently released and to the lack of a *Climate Change* section in the new Trump-era EPA website, which led the authors to select news releases as a further source of information. Nevertheless, it still appeared interesting to contrast the available data in order to investigate the 'transformation' of the US environmental science and strategies.

The texts were investigated with the support of the corpus analysis suite *WordSmith Tools 6.0* (Scott 2014) in order to identify recurring patterns. The corpus was also POS (part-of-speech) tagged with the aid of the online corpus query system *SketchEngine* (Kilgarriff *et al*. 2014) to establish word classes and syntactic categories. *SketchEngine* website also offers the function *Word Sketches*, i.e. corpus-derived summaries of the grammatical and collocational behaviour of words (Kilgarriff *et al*. 2010, p. 372), which may also be contrasted between subcorpora through the *Word Sketch Differences* utility.

In the initial phases of the study, a context-based analysis of the ecological discourse in the United States politics was performed via wider reading. The preliminary exploration highlighted a controversial rewriting and communication of climate change knowledge and policies under Trump's administration and guided corpus collection. Close reading of the corpus and computational analysis (e.g. of frequencies, clusters, keywords etc.) allowed to identify potential sites of interest in the documents along with possible discourses and strategies. Representative and significant sets of data within the corpus (namely, keywords, *word sketches*, concordances of key lexical items in the corpus and in a set of texts within the corpus) were analysed from a qualitative and CDA perspective, then contextualising findings in the wider social and political context (for further insights on the study framework combining CDA with corpus linguistics, see Baker *et al*. 2008, p. 295; Baker *et al*. 2013, p. 27). In particular, the present study compared Obama's and Trump's EPA websites, environmental reports, laws and information focusing on the selection, prioritisation and popularisation of scientific information, the quotation of authoritative sources to support positions and the exploitation of contrasting argumentations to justify political choices.

4. Analysis

4.1 The erased section on climate change

At the time of writing this paper, the EPA website is still being updated, however, the Obama-era version of the site is still available,[5] as it was migrated, frozen on January 19, 2017, i.e. the day before Trump's presidential inauguration. Comparing the two sites, in Trump's EPA website the issue of *climate change* does not appear to be listed any more among the main *Environmental Topics* (compare Figure 2 against Figure 1).

- Air
- Bed Bugs
- Chemicals and Toxics
- Climate Change
- Environmental Information by Location
- Greener Living
- Health
- Land, Waste, and Cleanup
- Lead
- Mold
- Pesticides
- Radon
- Water
- A to Z Index

Figure 1
EPA *Environmental Topics* during Obama's administration.

- Air
- Bed Bugs
- Chemicals and Toxics
- Environmental Information by Location
- Greener Living
- Health
- Land, Waste, and Cleanup
- Lead
- Mold
- Pesticides
- Radon
- Science
- Water
- A to Z Index

Figure 2
EPA *Environmental Topics* during Trump's administration.

Nonetheless, *climate change* is currently still present in the EPA alphabetical *Index of Environmental Topics* (see Figure 4). Yet, at present, the section contains limited materials and the main page on the theme in not clickable and accessible any more in the new version of the EPA website (compare Figure 4 against Figure 3).

[5] EPA (United States Environmental Protection Agency) 19.01.2017, *Snapshot*. https://19january2017snapshot.epa.gov/ (05.03.2018).

- Climate Change
  - Adaptation Resource Center (ARC-X), Climate Change
  - Analysis Report, Climate Change
  - Causes, Climate Change
  - Climate and Land-Use Scenarios
  - Climate Leadership Center
  - Effects, Climate Change
  - Facts, Climate Change
  - Indicators, Climate Change
  - Research, Climate Change
  - State, Local, and Tribal Governments, Climate and Energy Resources
  - Water Impacts, Climate Change

Figure 3
Climate Change in EPA *A-Z Index* during Obama's administration.

- Climate Change
  - Adaptation Resource Center (ARC-X), Climate Change
  - Analysis Report, Climate Change
  - Climate and Land-Use Scenarios
  - Climate Leadership Center
  - Indicators, Climate Change
  - Research, Climate Change
  - Water Impacts, Climate Change

Figure 4
Climate Change in EPA *A-Z Index* during Trump's administration.

When one tries to reach the general page on climate change on Trump's EPA website, one will be redirected to a notice page stating:

> This page is being updated.
> Thank you for your interest in this topic. We are currently updating our website to reflect EPA's priorities under the leadership of president Trump and Administrator Pruitt.[6]

Amazingly, as regards this matter, J.P. Freire, Associate Administrator for Public Affairs at the EPA, shamelessly justified the issue by declaring that the present agency was currently "removing outdated language":

> We want to eliminate confusion by removing outdated language first and making room to discuss how we're protecting the environment and human health by partnering with states and working within the law. (Freire in EPA 28.04.2017)

With a political act, Trump's administration is thus wiping out the most evident web pages on climate change. In plain words, Trump is now repudiating the language that was actually the result of years of internationally reviewed research. As the rewording process is still in progress, it will be particularly interesting to discover how scientific information will be further reshaped in the next future.

[6] EPA (United States Environmental Protection Agency), *Climate Change*. https://www.epa.gov/climatechange (10.03.2018).

4.1.1 Popularisation of scientific knowledge online

The Obama-era section on climate change, which has been completely erased from the current EPA website, aimed at providing information on the environmental issue to the general public, based on the most recent scientific data and exploiting several popularisation strategies (see e.g. Rasulo 2014). Causes and facts were, for instance, presented in the form of answers to the citizens' common doubts (see Figure 5) and readers were even engaged through a quiz testing their knowledge about the impacts of climate change[7].

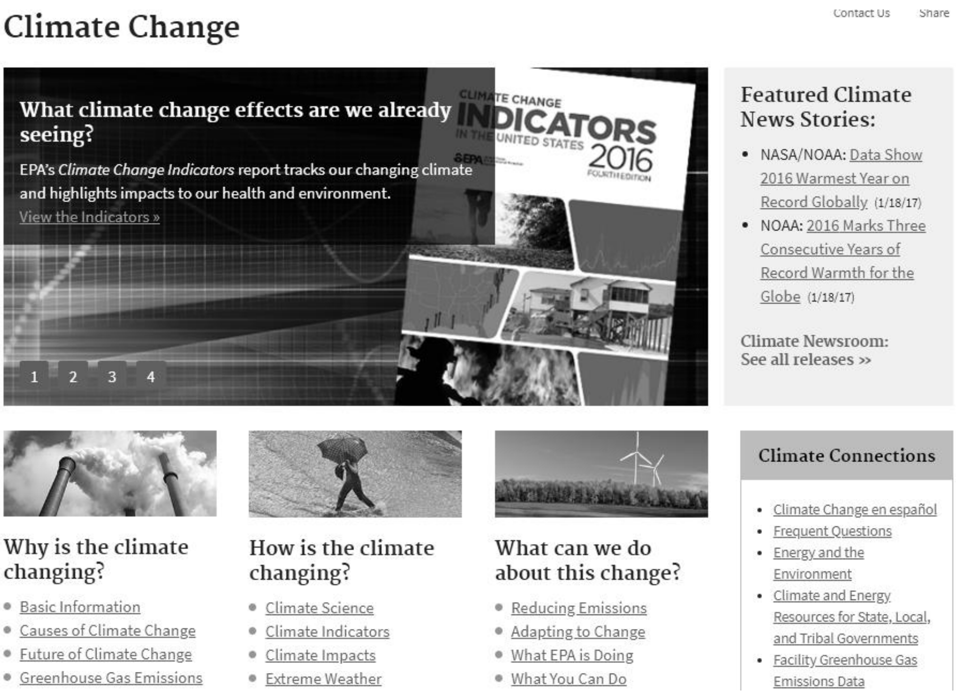
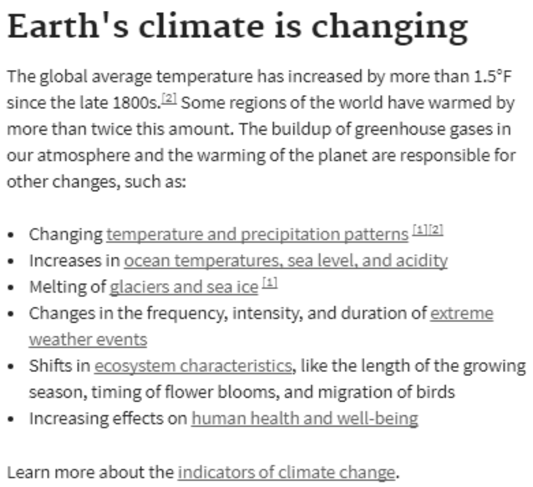


Figure 5
Climate Change page in EPA website during Obama's administration.

Climate change was not constructed as a remote hypothesis but as a fact, as shown by the statements in Figures 6 and 7. The impacts were exemplified through clear evidence, especially by listing the extreme weather conditions our planet is experiencing. As a common trend in the contemporary scientific discourse, ergative constructions, e.g. "Climate is changing", "our earth is warming", and nominalisations, such as "the buildup of greenhouse gases", "the warming of the planet", were used instead of transitive clauses to describe environmental processes taking place in the current world. As discussed by ecolinguistics experts (e.g. Alexander 1996; Gerbig 1993;

[7] EPA (United States Environmental Protection Agency) 19.01.2017, *Snapshot, Quiz: How Much Do You Know About the Health Impacts of Climate Change?* https://archive.epa.gov/epa/climate-impacts/text-version-quiz-how-much-do-you-know-about-health-impacts-climate-change.html (10.03.2018).

Goatly 2001), transitive constructions would explicitly identify the participants involved and clearly define the responsible actors and the affected participants in processes of environmental degradation and destruction. The use of ergative forms seems instead to frame the state of an entity as the result of some self-generating process (Goatly 2001, pp. 218-220). Agent deletion may also occur through the device of nominalisation, which puts less emphasis on the affected beings, suppressing agentivity, intention, motivation, and responsibility (Gerbig 1993; Goatly 2001; Schleppegrell 1997).



Earth's climate is changing

The global average temperature has increased by more than 1.5°F since the late 1800s.[2] Some regions of the world have warmed by more than twice this amount. The buildup of greenhouse gases in our atmosphere and the warming of the planet are responsible for other changes, such as:

- Changing temperature and precipitation patterns [1][2]
- Increases in ocean temperatures, sea level, and acidity
- Melting of glaciers and sea ice [1]
- Changes in the frequency, intensity, and duration of extreme weather events
- Shifts in ecosystem characteristics, like the length of the growing season, timing of flower blooms, and migration of birds
- Increasing effects on human health and well-being

Learn more about the indicators of climate change.

Figure 6
Climate Change page in EPA website during Obama's administration.

Climate change is happening

Our Earth is warming. Earth's average temperature has risen by 1.5°F over the past century, and is projected to rise another 0.5 to 8.6°F over the next hundred years. Small changes in the average temperature of the planet can translate to large and potentially dangerous shifts in climate and weather.

The evidence is clear. Rising global temperatures have been accompanied by changes in weather and climate. Many places have seen changes in rainfall, resulting in more floods, droughts, or intense rain, as well as more frequent and severe heat waves.

Figure 7
Climate Change page in EPA website during Obama's administration.

Nevertheless, in the environmental discourse of the EPA website under President Obama, lack of agentivity was not accompanied to a denial of human responsibility in environmental damage or preservation. Personalisation was instead even exploited to involve readers in the process of climate change mitigation, especially through the usage of the second person pronoun *you* (414 instances, 0.03% of the whole Obama subcorpus, against 32 occurrences, 0.01%, in the Trump corpus). In particular, the construction of *you* as a subject pronoun in the collocation *you can* (27 entries, 0.02%) appeared to be interestingly revealing. As noticeable from the concordance lines in Figure 8, the modal verb was used on the website especially when providing advice about what the common citizen *can do* at home, on the road and at work to reduce greenhouse gas emissions.

In Obama's discourse, the American people were thus not only presented as the receivers of political measures, but were also engaged as part of a responsible and forward-looking project to face one of the major challenges of the 21st century and preserve the natural world. The discourse of personal environmental commitment has instead been completely excised from the new EPA website. Such choice appears to be in line with Trump's tendency to remove any active requirements from citizens, representing the population as passively waiting to receive the due benefits from the government (see also Mettomäki 2017).

N	Concordance
1	What You Can Do: At Home Making a few small
2	you money. Explore our list of 10 simple steps you can take to reduce greenhouse gas
3	help you make the most energy-efficient decision. You can find the ENERGY STAR label on more
4	energy bill—about $1,000 a year! There is a lot you can do to drive down this cost. Simple steps
5	the environment. Depending on where you live, you can cut your annual energy bill by more than
6	the sun. There are two ways to use green power: You can buy green power, or you can modify
7	use green power: You can buy green power, or you can modify your house to generate your own
8	energy supply. There are a number of steps you can take to create a greener home,
9	]This rise in extreme weather events fits a pattern you can expect with a warming planet. Scientists
10	information about climate adaptation and things you can do to prepare for changes in extreme
11	Change Adaptation. For more information on how you can reduce greenhouse gas emissions at
12	, on the road, and in your workplace, see What You Can Do. References [1] USGCRP (2014).
13	emissions calculator, and much more. • What You Can Do about Climate Change on EPA's
14	Climate Change site offers suggestions for what you can do at home, at the office, at school, and
15	climate-related health risks by state and actions you can take to reduce these risks. • Air Quality
16	critical component of our atmosphere. However, you can have too much of a good thing. The
17	will live in. We can make a difference You can take action. You can take steps at
18	. We can make a difference You can take action. You can take steps at home, on the road, and in
19	or biking to work, can even improve your health! You can also get involved on a local or state
20	that can reduce these emissions, see What You Can Do. Other climate forcers Particles and
21	emissions, see What EPA is Doing and What You Can Do. Top of Page Changes in the sun's

Figure 8
Sample of concordance lines for *you can* in the Obama corpus.

The Obama EPA website summarised scientific knowledge about climate change also by endorsing the propositions with references to "highly credible and warrantable" (Martin, White 2005, p. 116) authoritative sources. In particular, as shown in Figure 9, expert studies were mentioned through the general noun *scientist** (173 entries, 0.01%, against only 20 occurrences in the Trump corpus, 0.006%, of which 19 in the NCA 2017 report and just 1 on the Trump-era EPA website, see Paragraph 4.5). Moreover, each online subsection on climate change also contained a set of bibliographical references citing official research, for a total of 72 quotations. The most quoted authorities were the EPA (7 mentions), the IPCC (7 quotes) and especially the US Global Change Research Program (USGCRP) (42 references), so that policies on environmental protection were construed as based on sound science.

N	Concordance
1	Weather Changes in Extreme Weather and Climate Events Scientists study many aspects of change in extreme
2	more erosion, threatening many communities with relocation. Scientists who study climate change confirm that these
3	certainty that climate change is happening now. While scientists continue to refine projections of the future,
4	from the top of the atmosphere to the depths of the oceans. Scientists and engineers from around the world have
5	in Greenland and Antarctica is limited, making it difficult for scientists to narrow the range of expected future sea level
6	factors have caused Earth's climate to change many times. Scientists have pieced together a record of Earth's climate,
7	of multiple factors, including natural variability. Nevertheless, scientists have been able to draw a connection between
8	the atmosphere to the depths of the oceans (Figure 2.1).3 Scientists and engineers from around the world have
9	." And around the same time we began exploring space, scientists were studying changes taking place in the Earth's
10	climate system (e.g., changes in ocean circulation patterns) Scientists use computer models of the climate system to
11	changes taking place in the Earth's atmosphere. Now, scientists had known since the 1800s that greenhouse
12	have repeatedly shown that more than 97 percent of scientists in the field agree that the world is unequivocally
13	of changes in average temperature or precipitation, climate scientists are actively researching the connections between
14	events fits a pattern you can expect with a warming planet. Scientists project that climate change will make some of
15	of both past and future emissions from human activities. Scientists are working to narrow the range of sea level rise
16	(see Figure 2.26 on global sea level rise).122,123 Although scientists cannot yet assign likelihood to any particular
17	agrees on every component of the climate change problem. Scientists are still researching a number of important
18	changes in oceans and clouds will affect climate change. Scientists continue to research these questions so society
19	), 1970 CAA Legis. Hist. at 1381. 206 For instance, while scientists, such as Stephen Schneider of the National
20	emissions and other human factors that influence climate, scientists use a range of scenarios using various
21	changes." 200 At that time, Congress heard the views of scientists that carbon dioxide emissions tended to increase

Figure 9
Sample of concordance lines for *scientist** in the Obama corpus.

4.2 The National Climate Assessments

The USGCRP is a US programme established in 1989 by Presidential Initiative to coordinate federal research on global environmental changes. The organisation is required to conduct a National Climate Assessment (NCA) every four years, resulting in a report to the President and the Congress. The document serves crucial functions, including: identifying advances in science, providing critical analysis of climate-related issues, highlighting key policy-relevant findings, guiding climate change decision-making.[8]

Under the second Obama's Presidency, the 2014 NCA report was published and, surprisingly, in spite of the current President's scepticism, on November 3, 2017, also the Trump administration released its 2017 NCA.

The 2014 document was preceded by a letter signed by Public Authorities (namely, the Assistant to the President for Science and Technology and by the National Oceanic and Atmospheric Administration Administrator), which clearly displayed President Obama's endorsement of the scientific findings divulged in the document. The letter stated that the report represented a strong base for governmental action, as it contained essential data which would guide Obama's climate policies and decisions:

[8] USGCRP (United States Global Change Research Program), *Legal Mandate*. https://www.globalchange.gov/about/legal-mandate (02.03.2018).

> [...] This information establishes a strong base that government at all levels of U.S. society can use in responding to the twin challenges of changing our policies to mitigate further climate change and preparing for the consequences of the climate changes that can no longer be avoided. It is also an important scientific resource to empower communities, businesses, citizens, and decision makers with information they need to prepare for and build resilience to the impacts of climate change.
> When President Obama launched his Climate Action Plan last year, he made clear that the essential information contained in this report would be used by the Executive Branch to underpin future policies and decisions to better understand and manage the risks of climate change. We strongly and respectfully urge others to do the same. [underlining added] (USGCRP 2014, p. iii)

On the contrary, Trump-era NCA did not seem to 'deserve' a similar approving introduction. In the colophon, the administration visibly distanced itself from the scientific research, insisting that the document had been published – only – to respond to a national requirement and that the law – not the Presidency – imposed it as a highly influential scientific assessment. A disclaimer was also added to stress that the report did not express any regulatory policies or "make any findings that could serve as predicates of regulatory action":

> This document responds to requirements of Section 106 of the U.S. Global Change Research Act of 1990 (P.L. 101-606, <http://www.globalchange.gov/about/legal-mandate>). It does not express any regulatory policies of the United States or any of its agencies, or make any findings of fact that could serve as predicates of regulatory action. Agencies must comply with required statutory and regulatory processes before they could rely on any statements in the document or by the USGCRP as basis for regulatory action.
> This document was prepared in compliance with Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (P.L. 106-554) and information quality guidelines issued by the Department of Commerce / National Oceanic and Atmospheric Administration pursuant to Section 515 (<http://www.cio.noaa.gov/services_programs/info_quality.html. For purposes of compliance with Section 515, this document is deemed a "highly influential scientific assessment" (HISA). [underlining added] (USGCRP 2017, p. iii)

Experts guaranteed that scientific contents were not altered in the 2017 report (Friedman, Thrush 03.11.2017). Nevertheless, if compared against the 2014 NCA, some rewordings and additions seem again to suggest that Trump's administration was forced to admit the existence of global warming (see also Mooney *et al.* 03.11.2017):

> Multiple lines of independent evidence confirm that human activities are the primary cause of the global warming of the past 50 years [underlining added] (USGCRP 2014, p. 7).

> Many lines of evidence demonstrate that it is extremely likely that human influence has been the dominant cause of the observed warming since the mid-20th century. Over the last century, there are no convincing alternative explanations supported by the extent of the observational evidence [underlining added] (USGCRP 2017, p. 14).

The 2014 NCA contained several references to President Barack Obama's programme *Climate Action Plan* (5 occurrences in the 2014 NCA), viewed as a positive mitigation act:

> *Adaptation activities in the United States* [...]
> • the release of President Obama's Climate Action Plan in June 2013, which has as one of its three major pillars, preparing the United States for the impacts of climate change, including building stronger and safer communities and infrastructure, protecting the economy and natural resources, and using sound science to manage climate impacts. (USGCRP 2014, p. 672)

No references to Donald Trump or to his energy policies were instead made in 2017 NCA. The President's worrying decision to withdraw from Paris Agreement was awkwardly mentioned in the most recent report (29 instances of *Paris* (*Agreement*)). Nevertheless, no comments were added on the consequences implied by this choice to rescind international commitments:

> In June 2017, the United States announced its intent to withdraw from the Paris Agreement. The scenarios assessed below were published prior to this announcement and therefore do not reflect the implications of this announcement (USGCRP 2017, pp. 397-398).

4.3 The environmental programmes

The present study also analyses comparatively the online documents condensing Obama's and Trump's environmental and energy plans.

In 2007, Obama launched his climate change mitigation programme, known as *Climate Action Plan*. The project rested on three pillars: cutting emissions while developing clean energy sources; building infrastructures to protect citizens from severe weather impacts resulting from climate change; acting not only at home but also maintaining international leadership in the fight against climate change:

> 1. Cut Carbon Pollution in America: [...] putting in place tough new rules to cut carbon pollution [...] and move our economy toward American-made clean energy sources that will create good jobs and lower home energy bills.
> 2. Prepare the United States for the Impacts of Climate Change: [...]

strengthen our roads, bridges, and shorelines so we can better protect people's homes, businesses and way of life from severe weather.

3. Lead International Efforts to Combat Global Climate Change and Prepare for its Impacts: [...] it is imperative for the United States to couple action at home with leadership internationally. America must help forge a truly global solution to this global challenge by galvanizing international action to significantly reduce emissions (particularly among the major emitting countries), prepare for climate impacts, and drive progress through the international negotiations. (Obama 2013, p. 5)

Trump's energy and environmental strategy was instead summarised in his *America First Energy Plan*, a name which echoed his election campaign slogan *America First*. The programme stands in diametrical opposition to the previous environmental policies, as it sets as its goals: first and foremost, eliminating Obama-era "harmful and unnecessary policies"; exploiting the national sources of energy, especially by boosting the coal industry; reaching energy independence; only "lastly", combining need for energy with a responsible protection of the environment.

> President Trump is committed to eliminating harmful and unnecessary policies such as the Climate Action Plan and the Waters of the U.S. rule […].
> We must take advantage of the estimated $50 trillion in untapped shale, oil, and natural gas reserves […].
> The Trump Administration is also committed to clean coal technology, and to reviving America's coal industry, which has been hurting for too long. […] President Trump is committed to achieving energy independence from the OPEC cartel and any nations hostile to our interests […].
> Lastly, our need for energy must go hand-in-hand with responsible stewardship of the environment.[9]

Challenging the regulations of Obama's era, Administrator Pruitt launched a *Back-To-Basics Agenda* for the EPA, which aims at subordinating environmental preoccupations to economic growth. Pruitt declared the "war on coal" was over (Pruitt 01.05.2017) and, with a deeply symbolical and populist act, he promoted the new energy plan among coal miners:

> [t]he coal industry was nearly devastated by years of regulatory overreach, but with new direction from President Trump, we are helping to turn things around for these miners and for many other hard working Americans. (Pruitt in EPA 13.04.2017)

Pruitt thus showed sympathy towards workers operating in the traditional energy industry, which represent a relevant part of Trump's pool of voters.

[9] The White House, *An America First Energy Plan*. https://www.whitehouse.gov/america-first-energy (30.10.2017).

The present analysis also considers the changes introduced in the EPA Strategic Plan, the document condensing the agency mission and goals, by comparing Obama's *14-18* programme against the *18-22* EPA agenda. The Obama-era EPA aims were summarised as follows:

1. Addressing Climate Change and Improving Air Quality
2. Protecting America's Waters
3. Cleaning Up Communities and Advancing Sustainable Development
4. Ensuring the Safety of Chemicals and Preventing Pollution
5. Protecting Human Health and the Environment by Enforcing Laws and Assuring Compliance (EPA 10.04.2014, p. 4)

In clear opposition to Obama's environmentalism, mentions to *climate change* and *sustainable development* were erased from Trump's EPA objectives:

1. Core Mission: Deliver real results to provide Americans with clean air, land, and water.
2. Cooperative Federalism: Rebalance the power between Washington and the states to create tangible environmental results for the American people.
3. Rule of Law and Process: Administer the law, as Congress intended, to refocus the Agency on its statutory obligations under the law. (EPA 02.2018, p. 4)

The new EPA mission insisted instead on bringing the agency back to its duties, its "statutory obligations under the law", with a clear reference to Pruitt's personal effort (see also Paragraph 1) and, in particular, to the political debate on the *Clean Power Plan* (CPP).

4.4 The Clean Power Plan and its Repeal

Promoted in 2015 by the Obama administration, the Clean Power Plan is a policy aimed at reducing carbon pollution by 32% by 2030 and incentivising the development of sustainable sources of energy. However, more than half of the American states and numerous industry groups are challenging the regulation, alleging that such controls over emissions exceed the EPA legal authority. The law was actually suspended by the U.S. Supreme Court, yet many states are continuing to support it and are already taking steps toward cleaner sources of energy, even without a mandate. During his campaign, Donald Trump had already criticised the law, and his administration recently issued a proposal of repeal, while promising to bring back coal mining jobs (see also Friedman, Plumer 09.10.2017).

As can be seen from Table 3, a subsection of the corpus under study collects the main official regulatory texts published on the EPA website on the issue, contrasting the documents establishing Obama's Clean Power Plan

(CPP) rule against the recent acts pursued by the Trump administration to advance its Repeal:

Obama's Clean Power Plan	**Trump's Repeal**
• *Final Rule. Clean Power Plan. Carbon Pollution Emission Guidelines for Existing Stationary Sources* (23.10.2015) • *Proposed Federal Plan for the Clean Power Plan* (23.10.2015) • *Final Rule. Carbon Pollution Standards for New, Modified and Reconstructed Power Plants* (23.10.2015)	• *Presidential Executive Order on Promoting Energy Independence and Economic Growth* (28.03.2017) • *Letter from Administrator Scott Pruitt to Kentucky Governor Matt Bevin re Clean Power Plan Guidance* (30.03.2017) • *Withdrawal of Proposed Rules Federal Plan Requirements for Greenhouse Gas Emissions from Electric Utility Generating Units Constructed on or Before January 8, 2014* (03.04.2017) • *Review of the Standards of Performance for Greenhouse Gas Emissions* (04.04.2017) • *Review of the Clean Power Plan* (04.04.2017) • *Memorandum Executive Order 13783 Promoting Energy Independence and Economic Growth* (19.04.2017) • *Proposal to Repeal the Clean Power Plan* (10.10.2017) • *Under Executive Order 13783 Final Report on Review of Agency Actions that Potentially Burden the Safe, Efficient Development of Domestic Energy Resources* (25.10.2017)

Table 3
Regulatory subsection of the corpus: Obama's CPP and Trump's Repeal acts.

In order to investigate the argumentations exploited under the two US Presidents to support political choices, the lists of the most frequent words for the CPP and Repeal subcorpora were obtained and compared with the aid of *WordSmith Tools* (Scott 2014) *Keywords* utility. The log likelihood test was used, setting the p value on 0.000001 and the minimum percentage of texts on 38%. Among the words characterising the two subsections, the analysis of two terms typifying Trump's Repeal corpus appeared to be particularly revealing: *burdens* and *benefits* (see Table 4).

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness
1	CPP	194	0,68	5	61		993,08
2	REVIEW	127	0,44	8	340	0,05	310,17
3	EXECUTIVE	91	0,32	7	129	0,02	306,52
4	AGENCY	138	0,48	8	535	0,07	260,93
5	ORDER	112	0,39	7	399	0,06	225,75
6	DOMESTIC	44	0,15	6	28		193,53
7	RESCIND	28	0,10	5	1		174,37
8	PRUITT	25	0,09	6	0		163,38
9	BENEFITS	85	0,30	5	411	0,06	133,14
10	SUSPEND	22	0,08	6	4		121,76
11	POLICY	47	0,16	8	138	0,02	108,21
12	REFORM	22	0,08	4	9		107,12
13	REGULATORY	77	0,27	7	421	0,06	107,08
14	EPA	401	1,40	7	5.700	0,79	106,05
15	S	246	0,86	8	2.950	0,41	103,09
16	PRESIDENT	28	0,10	7	34		100,26
17	LAW	45	0,16	7	152	0,02	94,19
18	STATUTORY	41	0,14	4	128	0,02	90,63
19	AUTHORITY	68	0,24	7	391	0,05	89,74
20	U	93	0,33	6	697	0,10	89,49
21	RECONSIDER	17	0,06	4	5		87,90
22	BURDENS	21	0,07	7	17		86,30

Table 4
Keywords: Repeal Corpus against CPP corpus.

Through *SketchEngine Sketch Differences* tool, the grammatical and collocational behaviour of the noun *burden** in the two subcorpora was identified (see Table 5). The lighter grey-coloured collocates represent those more peculiar to the Obama corpus, while the darker grey-coloured words are more typical of Trump's texts.

As evident from Table 5, the most frequent verb preceding *burden** is represented by *impose* in the Obama CPP section, and by *reduce* in Trump's Repeal files. The concordance lines of *impose* with *burden** in the CPP rule corpus were obtained, revealing that the Obama-era law tended to insist on the fact that no significant further regulatory and information collection burdens would be placed on power plants with the new emission rule, as in "will impose minimal new information collection burden" (see Figure 10).

From the concordances for the collocation of *reduce* with *burden** in the documents supporting the Repeal proposal (see Figure 11), it appeared instead clear that Trump's administration discourse aimed at conveying the exact opposite idea. Trump-era documents state that the CPP did create *unnecessary regulatory burdens*, which obstructed the development of domestic energy resources, while the new modifications will aim to *reduce* them, alleviating such encumbrance.

verbs with "burden" as object	48	27	0.53	1.00
minimize	7	0	11.0	--
bear	2	0	10.2	--
put	2	0	9.9	--
experience	2	0	9.5	--
relieve	1	0	9.3	--
define	3	0	9.2	--
place	1	0	8.8	--
allocate	1	0	8.1	--
present	1	0	7.7	--
add	1	0	7.4	--
increase	1	0	6.0	--
have	2	0	5.9	--
impose	10	2	10.8	8.6
report	3	1	8.1	6.6
estimate	2	1	8.5	7.7
associate	3	3	8.2	8.2
regard	2	2	8.1	8.2
create	1	1	7.3	7.4
reduce	2	9	6.2	8.4
review	0	1	--	8.0
avoid	0	3	--	9.0
remove	0	1	--	9.2
face	0	1	--	9.3
ease	0	1	--	9.8
alleviate	0	1	--	10.0

modifiers of "burden"	79	29	0.88	1.07
administrative	5	0	10.2	--
minimal	2	0	9.5	--
disproportionate	2	0	9.4	--
estimated	2	0	9.2	--
annual	5	0	8.9	--
respondent	1	0	8.7	--
undue	1	0	8.7	--
lesser	1	0	8.6	--
reporting	2	0	8.4	--
great	2	0	8.1	--
corresponding	1	0	7.9	--
financial	1	0	7.9	--
less	1	0	7.7	--
direct	1	0	7.6	--
cost	2	0	7.6	--
environmental	2	0	7.4	--
substantial	1	0	7.4	--
reduction	2	0	7.4	--
significant	2	0	7.3	--
information	10	1	10.6	7.7
collection	10	1	11.5	9.0
unnecessary	5	11	10.7	12.9
regulatory	3	12	7.9	10.1
tremendous	0	1	--	10.0
potentially	0	3	--	11.5

Table 5
Sketch Differences: collocates of *burden** in the CPP corpus (lighter grey) against the Repeal Corpus (darker grey).

captured carbon from an EGU due to the reporting and other regulatory	burdens	*imposed* by the monitoring requirements of GHGRP subpart RR.484 They
is acceptable, and use of the technology does not *impose* significant	burdens	on energy requirements at either the plant or national level. The 1,400
EGUs This final action will *impose* minimal new information collection	burden	on owners and operators of affected newly constructed fossil
This final action is not expected to *impose* an information collection	burden	under the provisions of the PRA on owners and operators of affected
EGUs are not anticipated to *impose* any information collection	burden	over the 3-year period covered by this ICR. We have estimated, however,
, this final action would *impose* minimal information collection	burden	on those affected EGUs beyond what they would already be subject to
in building block 2 would not *impose* significant additional	burden	on the transmission planning process and does not necessitate major
and, as a result, is not projected to *impose* a substantial cost	burden	on affected facilities. With respect to entrainment, the rule calls
, this action is not anticipated to *impose* any information collection	burden	on tribal governments over the 3-year period covered by this ICR. This
state having to submit a new SIP. This limitation *imposed* unnecessary	burdens	on states to fix even obvious errors, because CAA section 110(a)(2)

Figure 10
Sample of concordance lines for *impose* + *burden** in the Obama CPP section.

the development of U.S. energy resources and to *reduce* unnecessary regulatory burdens associated with the development of those resources. The President has directed
energy resources-including oil and gas-and to *reduce* unnecessary regulatory burdens associated with the development of those resources. The President has directed
the development of U.S. energy resources and to *reduce* unnecessary regulatory burdens associated with the development of those resources. The President has directed
safe development of our energy resources while *reducing* unnecessary regulatory burdens and directs federal agencies to undertake several actions to further this goal.
this review that includes recommendations for *reducing* unnecessary regulatory burdens . Through implementation of environmental statutes such as the Clean Air Act and
key initiatives that it believes will further the goal of *reducing* unnecessary burdens on the development and use of domestic energy resources. These initiatives
to re-examine how it engages with industry to *reduce* unnecessary regulatory burdens , improve regulatory forecasting and predictability, and improve the ability of
identified herein will advance the goal of *reducing* unnecessary regulatory burdens on the development and use of domestic energy resources in accordance with E.O.
the information that the Agency was collecting through these requests, and *reduce* burdens on businesses while the Agency assesses such need. EPA estimated the burden of the

Figure 11
Sample of concordance lines for *reduce* + *burden** in the Trump Repeal section.

The collocation of the term *benefit** was also analysed in the two subcorpora. In particular, the pre-modifiers of the noun were considered, since revealing of the major areas recipient of positive impacts. Table 6 lists the most frequent L1 modifiers of *benefit** in Obama’s Clean Power Plan and in Trump’s Repeal subcorpora, specifying for each word their absolute and normalised frequency in the subsection.

Obama’s Clean Power Plan			**Trump’s Repeal**		
Word	**Freq.**	**%**	**Word**	**Freq.**	**%**
climate	69	0.009%	forgone	23	0.075%
monetized	30	0.004%	net	16	0.052%
net	29	0.004%	climate	7	0.023%
health	23	0.003%	PM 2.5	6	0.020%
environmental	18	0.002%	air quality	5	0.016%
additional	10	0.001%	efficiency	4	0.013%
ecosystem	10	0.001%	health	1	0.003%
non-monetized	10	0.001%			
overall	8	0.001%			
CO2	6	0.001%			
economic	5	0.001%			
quantified	5	0.001%			
broad	4	0.001%			
carbon	4	0.001%			
social	3	0.000%			

Table 6
L1 modifiers of *benefit**.

It can be observed that, in the Obama corpus, *benefit** is mainly preceded by words referring to the positive impacts of the environmental law on nature (e.g. *climate*), people (e.g. *health*) and economy (e.g. *monetized*). Such advantages appear instead to be clearly understated in the Trump corpus, as suggested by the frequent collocation *forgone benefits*.

In particular, the text of the Repeal Proposal undervalued the relevance of mitigation acts, also by claiming that the actual benefits of the Clean Power Plan were still “highly uncertain”. Conversely, the Trump administration capitalised on the fear that emission regulations may instead damage the industry, thus leading to job loss. This would carry devastating results, especially in areas with limited re-employment opportunities. With a skilful rhetoric twist, the role of environment on human health was subordinated to occupational concerns, since "job loss may increase risks to health, of substance abuse, and even of mortality":

> With respect to the forgone benefits associated with this action, the EPA conducted a proximity analysis for the CPP which showed a higher percentage of <u>low-income and minority households living in proximity to EGUs that may have reduced emissions under the CPP</u>. These communities may experience forgone benefits as a result of this action. However, any changes in ambient air quality depends on stack height, atmospheric conditions, and dispersion patterns. Therefore, <u>the distribution of forgone benefits is highly uncertain</u>. […] Workers losing jobs in regions or occupations with weak labor markets would have been most vulnerable. With <u>limited re-employment opportunities</u>, or if new employment offered lower earnings, then unemployed workers could face extended periods without work, or permanently reduced future earnings. In addition, past research has suggested that involuntary <u>job loss may increase risks to health, of substance abuse, and even of mortality</u>. These adverse impacts may be avoided with the proposed repeal of the CPP. [underlining added] (EPA 10.10.2017)

4.5 The EPA news releases

The Trump administration's rewriting of the American environmental discourse also emerged from the analysis of the EPA new news releases about Climate and Energy policies. Through *WordSmith Tools* (Scott 2014) *WordList* function, the most frequent words in the reports were obtained. After having discarded function words and proper names, the corpus showed up four lexical groups of terms, referring to four main semantic areas: regulations, economy, energy and environment:

- Regulations: e.g. *executive order* (74 occurrences in the News Reports section, with a normalised frequency in the news releases subcorpus of 0.58%), *rule* (68, 0.54%), *regulation* (49, 0.39%), *Clean Power Plan* (44, 0.35%), *regulatory* (34, 0.27%), *repeal* (17, 0.13%), *Paris* (14, 0.11%).
- Energy: e.g. *energy* (70, 0.55%), *coal* (61, 0.48%), *gas* (32, 0.25%), *industry* (25, 0.2%), *resource* (17, 0.13%), *oil* (18, 0.14%), *electric* (15, 0.12%), *fuel* (15, 0.12%).
- Economy: *job* (35, 0.28%), *economy* (34, 0.27%), *cost* (23, 0.18%), *economic* (20, 0.16%), *business* (20, 0.16%), *miner* (15, 0.12%).
- Environment: e.g. *environment* (45, 0.35%), *protect* (27, 0.21%), *protection* (24, 0.19%), *emission* (22, 0.17%), *air* (21, 0.17%), *climate* (18, 0.14%), *water* (11, 0.09%), *greenhouse* (10, 0.08%).

Among the most frequent terms in the Trump news reports, it appeared interesting to consider the concordances for the term *environment* itself (see Figure 12).

on its intended mission, returning power to the states, and *creating*	an environment	where jobs can grow. "What better way to launch EPA's Back-to-Basics
'm committed to working in coordination with states to *create* a healthy	environment	where jobs and businesses can grow. That's the purpose of my
and local communities as we develop a path forward to *improve* our	environment	and bolster the economy in a manner that is respectful of and consistent
the capital that is spent on innovating new technologies to *improve* our	environment	even more. U.S. Chamber's Donohue Praises President Trump's
make sure that we transform our economy, grow jobs and also *protect* our	environment	. And it's an exciting day and I appreciate his leadership. I now have the
states' role and reinstating transparency into how we *protect* our	environment	," said Administrator Pruitt. Background On March 28, President Trump
, good growth and pro-energy policies at the same time as *protecting* our	environment	. The president is sending that message today, by his signature today on
, Washington has sold a false choice to the American people: *choose* the	environment	or jobs. That false choice today gives way to a 'show-and-tell' story-
Administrator Pruitt to develop sensible policies that *protect* the	environment	without shutting down more coal-fueled power plants, one of our most
will continue to review these actions to ensure that they *protect* the	environment	and enable a growing economy. "EPA is continuing to ensure that the
alleged "Clean Power Plan" would've done virtually zero to *protect* the	environment	while at the same time destroying thousands of jobs and costing
gives way to a 'show-and-tell' story-about our ability to *protect* the	environment	and create jobs simultaneously-that manufacturers are eager to lead.
means returning EPA to its core mission: *protecting* the	environment	by engaging with state, local, and tribal partners to create sensible
, which aims to refocus the agency on its core mission of *protecting* the	environment	through sensible regulations developed in cooperation with state,
agenda. The Back-to-Basics agenda refocuses the EPA on *protecting* the	environment	, promoting economic and job growth, and returning power to the states.
language first and making room to discuss how we're *protecting* the	environment	and human health by partnering with states and working within the law."
regulatory certainty to American job creators while *protecting* the	environment	. Learn more here. News Releases from Headquarters EPA to Reconsider
We should not have to choose between supporting jobs and *supporting* the	environment	. Now, opponents of President Trump's new executive order claim that

Figure 12
Sample of concordance lines for verbs with *environment* as object in Trump News Reports section.

The close reading of the concordance lines about environmental stewardship (e.g. *environment* in collocation with *protect*, *improve*, *support*) reveals the exploitation of one of the President’s typical discourses, part of his populist political agenda. As exemplified in his Inaugural Address (Trump 20.01.2017), Donald Trump tends to construe a bleak current scenario caused by his predecessor, which he contrasts with the brighter future he will provide to the nation (see also Napolitano 2018). Similarly, concerning environmental and energy policies, Trump-era texts stress that, while Obama had chosen global environment over American jobs, the current Presidency will instead combine environmental safeguard with economic growth.

A rhetoric similar to the one identified for the term *environment* appears to be exploited in Trump’s EPA news reports when another key term of environmental discourse, the noun *climate*, is used. As evident from Figure 13, such word does not seem to express involvement for natural concerns but to convey the idea that previous emission cuts under the Clean Power Plan and Paris Agreement were harmful for the US economy, especially for the coal sector, while showing no meaningful impact on the environment. The present administration is instead presented as acting in the national interest by revoking such regulations.

energy executive order that will restore sanity after Obama's failed	climate	policies. "I commend President Trump for sending this flawed
coal industry applauded an executive order starting a roll back of	climate	regulation signed by President Donald Trump on Tuesday. Lignite
in place, these regulations would have little positive impact on our	climate	while continuing to decimate West Virginia's economy," U.S. Rep.
. The previous administration, with every step they took in their	climate	action agenda, still fell 40% short of those 26 to 28% targets. What the
PRUITT: "When the Paris Accord was cut, the head, father of	climate	science, James Hanson, former NASA scientist, called the Paris Accord
on two and half trillion dollars in aid that would pass through the green	climate	fund. . . ... This whole process in Paris was all about talk. The previous
administration's analysis of CPP include: Domestic versus global	climate	benefits: The previous administration compared U.S. costs to an
on Thursday, April 13. A photo ID is required. R056 United States Resets	Climate	Change Discussion At G7 U.S. Formally Joins Communiqué, Reaching
withdraw from the Paris Agreement and has reset the conversation about	climate	change reflective of the new priorities of the Trump Administration
allies and key international partners, we approached the	climate	discussions head on from a position of strength and clarity. We are
. While a party to the communiqué, the United States did not join the	climate	change sections, explicitly stating: We the United States of America
, we the United States do not join those sections of the communiqué on	climate	and MDBs, reflecting our recent announcement to withdraw and
, and European Commissioners responsible for environment and	climate	, met in Bologna on 11-12 June 2017. We were joined by heads and senior
Order rescinds several Obama executive orders and policies related to	climate	change. It also directs the Administrator of the Environmental
the last administration, is out of date. Similarly, content related to	climate	and regulation is also being reviewed. News Releases from
West Virginia's vital coal industry while having no meaningful	climate	impact. Stopping this disastrous plan will preserve America's coal
, who would be hardest hit by the previous administration's harmful	climate	regulations. Americans depend on affordable, reliable electricity
the process to unwind the CPP, the Obama administration's signature	climate	change regulation that was stayed by the Supreme Court one year ago.

Figure 13
Sample of concordance lines for *climate* in Trump News Reports section.

The news reports analysed often contain portions of Administrator Pruitt's public interventions. Incredibly enough, in his rhetorical justification supporting the annulment of environmental commitments, Pruitt even (mis)quoted an authoritative source. Pruitt said James Hansen, the "father of climate science", called the Paris Agreement a 'fake' and a 'fraud':

> When the Paris Accord was cut, the head, father of climate science, James Hansen, former NASA scientist, called the Paris Accord a fake and a fraud. (Pruitt 06.06.2017)

Hansen's words would thus seem to echo a 'Trumpist' language, a support for the new President's policies and climate denial discourse (see also Paragraph 1). Yet, Pruitt actually conveyed this external opinion in a distorted way. As a matter of fact, the scientist had not opposed the mitigating policies but only the slow pace of Paris emission cuts. Against Trump, Hansen even acknowledged his government was "in the pocket of the fossil fuel industry":

> Governments that say climate change is a problem and then propose half-baked solutions that don't solve anything are in some ways a bigger problem than the Trump-type governments. With those, everybody got to see what they were doing and that they were in the pocket of the fossil fuel industry. (Hansen in Özbek 10.11.2017)

5. Preliminary conclusions

Climate change constitutes one of the major challenges of our time, requiring urgent action, especially due to the fast pace and global scale by which greenhouse gases are accumulating in the atmosphere. The United States, one of the world's major emitters, had recently shown a particular interest towards sustainable development and climate change mitigation. Regrettably, the current US President is overturning much of the past progress. Donald Trump is exploiting a climate sceptical attitude to promote a business-oriented political agenda, based on the idea that environmental concerns have led previous governments to neglect human needs, while the present administration will put American interest before environmentalist extremisms.

The present study analysed a diversified collection of environmental documents published on institutional US websites, comparing texts from the Obama's era with those produced during Trump's administration. It explored the publications through a framework integrating CDA and corpus linguistics, focusing on how scientific information was selected and communicated, which authoritative sources were quoted and which arguments were exploited to sustain political decisions.

Obama's environmental discourse displayed a clear and continued effort towards sustainable development. The former President showed respect and support for acknowledged international research, which was proficiently popularised for the general public, trying to engage Americans in sustainable behaviours. Scientific knowledge represented a guide for concrete policies aimed at reducing greenhouse gas emissions, revealing a commitment for the domestic fight against climate change at home and a wish to transfer the traditional American appeal for international leadership also to the environmental field.

The texts produced under Trump's administration reveal instead a strong aversion and opposition to Obama's policies. The current President discredits the importance of ecological preoccupations and shows a disdainful underestimation of climate issues. Scientific integrity is being abused and relevant information is being cancelled from official websites. The most recent documents display the President as trying to manage the US like his own company, following a 'shareholder approach' for which the only social responsibility of a business seems to be "to increase its profits" (Friedman 13.09.1970). Trump appears to consider the environment as functional to economic goals and nature as a source of goods, not a resource to be preserved for future generations (see also Stibbe 2014). The rhetoric of *America first* and *Back-to-Basics* 'no-frills' plans would thus leave no space to the expression of environmental values. The communications analysed seem to construe an ill antithesis between the environment and the people. An anachronistically irresponsive exploitation of nature seems thus to be justifiable if the purpose is to create jobs, overcome the crisis and promote popular well-being.

The present study represented a preliminary research on the rewriting of the US climate discourse. At the time of writing, governmental websites are still being updated and adapted to Trump's ideology and priorities. It would thus appear interesting to expand the analysis by considering the new developments and changes enacted by Trump which are expected to further dismantle the US environmental discourse and policies.

Bionotes: Antonella Napolitano is Researcher in English Language and Translation at the University of Sannio, Benevento, Department of Law, Economics, Management and Quantitative Methods. Her research activity has centred on the field of ESP theory and applications, professional and institutional discourse, the language of advertising, legal English, identity-building and gender studies. Some of her recent works focus on online customer reviews, corporate scandals and the discourse of and about President Trump. She is currently conducting a research on Institutional and corporate communication and new media. She is also scientific coordinator of the Language Centre at the University of Sannio.
Maria Cristina Aiezza is Post-Doctoral Research Fellow and Lecturer in English Language and Translation at the University of Sannio, Benevento, Department of Law, Economics, Management and Quantitative Methods. She holds a PhD in ESP from the University of Naples Federico II, with a corpus-assisted study on the discourse of CSR reports. Her research interests and publications focus on corporate discourse, institutional and political discourse, legal discourse, environmental discourse, popularisation, discourse of news media, web genres and social media. She is currently collaborating with Antonella Napolitano on a research project on Institutional and corporate communication and new media.

Authors' addresses: antonella.napolitano@unisannio.it; mariacristina.aiezza@unisannio.it

References

Alexander R.J. 1996, *Integrating the Ecological Perspective: Some Linguistic Self Reflexions*, in Fill A. (ed.), *Sprachökologie und Ökolinguistik*, Stauffenburg, Tübingen, pp. 131-148.

Alexander R.J. and Stibbe A. 2013, *From the Analysis of Ecological Discourse to the Ecological Analysis of Discourse*, in "Language Sciences" 41, pp.104-110.

Anthony A. 18.03.2017, *The Climate Change Battle Dividing Trump's America*, *The Guardian*. https://www.theguardian.com/science/2017/mar/18/the-scientists-taking-the-fight-to-trump-climate-change-epa (08.03.2018).

Baker P. 2006, *Using Corpora in Discourse Analysis*, Continuum, London/New York.

Baker P., Gabrielatos C., Khosravinik M., Krzyzanowski M., McEnery T., and Wodak R. 2008, *A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees And Asylum Seekers in the UK Press*, in "Discourse and Society" 19, pp. 273-306.

Baker P., Gabrielatos C. and McEnery T. 2013, *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*, Cambridge University Press, Cambridge.

Bhatia V.K. 2004, *Worlds of Written Discourse: A Genre-Based View*, Continuum, London/New York.

Boden T.A., Marland G. and Andres R.J. 2017, *Global CO_2 Emissions from Fossil-Fuel Burning, Cement Manufacture, and Gas Flaring, 1751-2014*, Oak Ridge, USA.

EDGI (Environmental Data and Governance Initiative), *Website Monitoring*. https://envirodatagov.org/website-monitoring/ (02.03.2018).

EPA (United States Environmental Protection Agency). https://www.epa.gov/ (02.03.2018).

EPA (United States Environmental Protection Agency) 10.04.2014, *Fiscal Year 2014-2018 EPA Strategic Plan*, EPA. https://nepis.epa.gov/Exe/ZyPDF.cgi/P100KB1L.PDF?Dockey=P100KB1L.PDF (10.03.2018).

EPA (United States Environmental Protection Agency) 02.2018, *FY 2018-2022 EPA Strategic Plan*, EPA. https://www.epa.gov/sites/production/files/2018-02/documents/fy-2018-2022-epa-strategic-plan.pdf (10.03.2018).

EPA (United States Environmental Protection Agency) 13.04.2017, *EPA Launches Back-To-Basics Agenda at Pennsylvania Coal Mine*, EPA News Releases. https://www.epa.gov/newsreleases/epa-launches-back-basics-agenda-pennsylvania-coal-mine (02.03.2018).

EPA (United States Environmental Protection Agency) 28.04.2017, *EPA Kicks Off Website Updates*, EPA News Releases. https://www.epa.gov/newsreleases/epa-kicks-website-updates (02.03.2018).

EPA (United States Environmental Protection Agency) 10.10.2017, *40 CFR Part 60, Proposed Rule: Repeal of Carbon Pollution Emission Guidelines for Existing Stationary Sources: Electric Utility Generating Units*. https://www.gpo.gov/fdsys/pkg/FR-2017-10-16/pdf/2017-22349.pdf (02.03.2018).

Fairclough N. 1992, *Discourse and Social Change*, Polity Press, Cambridge, UK/ Malden, USA.

Fairclough N. 1995, *Media Discourse*, Arnold, London.

Fairclough N. 2004, *Critical Discourse Analysis in Researching Language in the New*

Capitalism: Overdetermination, Transdisciplinary and Textual Analysis, in Harrison C. and Young L. (eds), *Systemic Linguistics and Critical Discourse Analysis*, Continuum, London, pp. 103-22.

Friedman M. 13.09.1970, *The Social Responsibility of Business is to Increase its Profits*, *The New York Times Magazine*. http://www.nytimes.com/1970/09/13/archives/a-friedman-doctrine-the-social-responsibility-of-business-is-to.html (02.03.2018).

Friedman L. and Plumer P. 09.10.2017, *E.P.A. Announces Repeal of Major Obama-Era Carbon Emissions Rule*, *The New York Times*. https://www.nytimes.com/2017/10/09/climate/clean-power-plan.html (10.02.2018).

Friedman L. and Thrush G. 03.11.2017, *U.S. Report Says Humans Cause Climate Change, Contradicting Top Trump Officials*, *The New York Times*. https://www.nytimes.com/2017/11/03/climate/us-climate-report.html (02.03.2018).

Gerbig A. 1993, *The Representation of Agency and Control in Texts on the Environment*, in Alexander R.J., Bang J.C. and Døør J. (eds), *Papers for the Symposium "Ecolinguistics, Problems, Theories and Methods", AILA 1993*, Odense University, Odense, pp. 61-73.

Goatly A. 2001, *Green Grammar and Grammatical Metaphor, or Language and Myth of Power, or Metaphors We Die By*, in Fill A. and Mühlhäusler P. (eds), *The Ecolinguistics Reader: Language, Ecology and Environment*, Continuum, London/New York, pp. 203-225.

Gotti M. 2005, *Investigating Specialized Language*, Peter Lang, Bern.

Greshko M., Parker L. and Howard B. 25.10.2017, *A Running List of How Trump is Changing the Environment*, *The National Geographic*. https://news.nationalgeographic.com/2017/03/how-trump-is-changing-science-environment/ (02.03.2018).

Grundmann R. and Krishnamurty R. 2010, *The Discourse of Climate Change: A Corpus-Based Approach*, in "CADAAD (Critical Approaches to Discourse Analysis across Disciplines" 4 [2], pp. 113-133.

IPCC (United Nations Intergovernmental Panel on Climate Change) 2013, *Climate Change 2013. The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, New York.

Kilgarriff A., Kovář V., Krek S., Srdanovic I. and Tiberius C. 2010, *A Quantitative Evaluation of Word Sketches*, in Dykstra A. and Schoonheim, T. (eds), *Proceedings of the 16th EURALEX (European Association for Lexicography) International Congress*, Fryske Academy, Leeuwarden, The Netherlands, pp. 372-379.

Kilgarriff A., Baisa V., Bušta J., Jakubíček M., Kovvář V., Michelfeit J., Rychlý P. and Suchomel V. 2014, *The Sketch Engine: Ten Years on*, in "Lexicography" 1, pp. 7-36.

Luntz F. 2002, *Straight Talk: Memorandum to Bush White House by GOP Consultant Frank Luntz*. https://www.sourcewatch.org/images/4/45/LuntzResearch.Memo.pdf (02.03.2018).

Martin J.R. and White P.R.R. 2005, *The Language of Evaluation: Appraisal in English*, Palgrave Macmillan, New York.

Matthews D. 01.06.2017, *Donald Trump Has Tweeted Climate Change Skepticism 115 Times: Here's all of it*, *The Vox*. https://www.vox.com/policy-and-politics/2017/6/1/15726472/trump-tweets-global-warming-paris-climate-agreement (09.03.2018).

Mettomäki A. 2017, *'Yes We Can' or 'Make America Great Again': Comparison of*

Representations in Barack Obama and Donald Trump's Inaugural Addresses, Unpublished Bachelor's Thesis, University of Jyväskylä, Department of Language and Communication Studies, Jyväskylä, Finland.

Milman O. and Rushe D. 22.02.2017, *New EPA Head Scott Pruitt's Emails Reveal Close Ties with Fossil Fuel Interests*, *The Guardian*. https://www.theguardian.com/environment/2017/feb/22/scott-pruitt-emails-oklahoma-fossil-fuels-koch-brothers (09.03.2018).

Milman O. and Morris S. 14.05.2017, *Trump is Deleting Climate Change, One Site at a Time*, *The Guardian*. https://www.theguardian.com/us-news/2017/may/14/donald-trump-climate-change-mentions-government-websites (02.03.2018).

Mooney C., Eilperin J. and Dennis B. 03.11.2017, *Trump Administration Releases Report Finding 'No Convincing Alternative Explanation' for Climate Change*, *The Washington Post*. https://www.washingtonpost.com/news/energy-environment/wp/2017/11/03/trump-administration-releases-report-finds-no-convincing-alternative-explanation-for-climate-change/?utm_term=.f71f9e89aeb7 (02.03.2018).

Napolitano A., 2018, *Another Brick in the Wall. A Critical Discourse Analysis of the Political Discourse of Division*, in Diglio C., Napolitano A. and Perilli F. (eds), *Identity, Language and Diversity. Between Walls and Bridges*, Paolo Loffredo Iniziative Editoriali, Naples, pp. 469-494.

Obama B. 2013, *The President's Climate Action Plan*, *The White House*. https://obamawhitehouse.archives.gov/sites/default/files/image/president27sclimateactionplan.pdf (10.03.2018).

Özbek C. 10.11.2017, *COP23: Famed Climate Scientist Hansen Warns of 'Bigger Problems' than Trump*, *Deutsche Welle*. http://www.dw.com/en/cop23-famed-climate-scientist-hansen-warns-of-bigger-problems-than-trump/a-41315823 (02.03.2018).

Pruitt S. 01.05.2017, *EPA is Putting American Workers First*, *The Washington Times*. https://www.washingtontimes.com/news/2017/may/1/epa-is-putting-american-workers-first/ (02.03.2018).

Pruitt S. 06.06.2017, *Pruitt: President Trump Is Focused On Growing Our Economy*, *EPA News Releases*. https://www.epa.gov/newsreleases/pruitt-president-trump-focused-growing-our-economy (10.03.2018).

Rasulo M. 2014, *Popularizing the CLIL Classroom*, in "ESP Across Cultures" 11, pp. 121-135.

Schleppegrell M.J. 1997, *Agency in Environmental Education*, in "Linguistics and Education" 9 [1], pp. 49-67.

Scott M. 2014, *WordSmith Tools 6.0*, Oxford University Press, Oxford.

Stibbe A. 2014, *Ecolinguistics and Erasure: Restoring the Natural World to Consciousness*, in Hart C. and Cap P. (eds), *Contemporary Critical Discourse Studies*, Bloomsbury Academic, London, pp. 583-602.

Trump D. 20.01.2017, *The Inaugural Address*, *WhiteHouse.gov*. https://www.whitehouse.gov/briefings-statements/the-inaugural-address/ (09.03.2018).

Trump D. 01.06.2017, *Statement by President Trump on the Paris Climate Accord*, *The White House*. https://www.whitehouse.gov/briefings-statements/statement-president-trump-paris-climate-accord/ (09.03.2018).

UN (United Nations) 1992, *United Nations Framework Convention on Climate Change*, United Nations, New York.

USGCRP (United States Global Change Research Program). https://www.globalchange.gov/ (02.03.2018).
USGCRP (United States Global Change Research Program) 2014, *Climate Change Impacts in the United States: The Third National Climate Assessment*, US Global Change Research Program.
USGCRP (United States Global Change Research Program) 2017, *Climate Science Special Report. Fourth National Climate Assessment Volume I*, US Global Change Research Program, Washington, DC, USA.
Van Dijk T. 2011, *Discourse Studies: A Multidisciplinary Introduction. Second Edition*, Sage, Thousand Oaks, USA.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 179-204
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p179
http://siba-ese.unisalento.it, © 2019 Università del Salento

DISSEMINATING CLIMATE CHANGE KNOWLEDGE
Representation of the International Panel on Climate Change in three types of specialized discourse

CAMILLE BIROS, CAROLINE PEYNAUD
UNIVERSITÉ GRENOBLE ALPES

Abstract – Regarding such a complex issue as climate change, it is crucial to understand how specialised knowledge is conveyed to the public. The Intergovernmental Panel on Climate Change (IPCC) is often considered the main source of scientific knowledge on climate change. This paper aims at understanding how knowledge produced by the IPCC is disseminated in three types of documents, with various degrees of specialisation, Earth Negotiation Bulletins, United Nations reports and the general-interest press, over two time periods (2009-2010 and 2014-2017). To do so, the analysis combines the tools of corpus linguistics and discourse analysis, which puts quantitative results in context. Identifying IPCC collocates enables to specify the context of use according to sub-corpora. This reveals that although the IPCC is referred to as a knowledge-provider in all corpora, there are variations as to the degree of trustworthiness given to the institution as well as to the amount of detail given on results. Looking at explicit quotations according to their theme, wording and type of reported speech confirms this. Recontextualization processes in the press corpora entail more reformulation and explanatory comments. The scientific nature of the work is made more explicit in the United Nations reports and Earth Negotiation Bulletins. Differences between the six sub-corpora can be analysed according to generic specificities and time frames revealing the increasingly central role of the IPCC as a trustworthy provider of scientific knowledge on climate change.

Keywords: International Panel on Climate Change; Specialised Knowledge; Recontextualisation; Genre Analysis.

1. Introduction

To assess human-induced climate change and its effects, it is necessary to draw on knowledge from diverse disciplines, such as, for example natural physics, climatology, geography, economics, political theory, to name but a few (Colson *et al.* 2009). This helps explain the importance of the role of the International Panel on Climate Change (IPCC) in publishing referential reports on human-induced climate change. Although this organisation does not produce new results on the topic, it is quoted as a reference on climate change knowledge owing to its function of synthetizing and evaluating all available publications on this topic. Given the number of disciplines involved, knowledge on climate change is not driven by disciplinary specialization but by policy needs, as the international community strives to find applicable solutions against climate change. The IPCC responds to these needs by publishing comprehensive research on the subject as well as giving advice to policymakers. On account of this role and its link to intergovernmental decision-making organisations, the status of the specialized knowledge it provides may be a topic of controversy (Aykut, Dahan 2015). Our aim here is to examine this role by looking at how the content of its reports is disseminated through different forms of communication. To do so we constituted a corpus of published, written documents on climate change that can be broken down into six sub-corpora reflecting generic specificities and diachronic variation. We intend to find out whether the status of the IPCC as a specialized knowledge-provider is context-specific. To do so, we consider how the content of the reports is disseminated across different genres and time periods. After a brief presentation of the theoretical framework to our analysis, we use corpus linguistic tools to analyse the contexts in which the IPCC's name is mentioned. Collocates and concordance lines of the term IPCC are analysed, and quotations of the IPCC are considered in context and classified according to content, type of quotation and wording. Our overall purpose is to contribute to a better understanding of the role of the IPCC as a knowledge-provider and to track discursive variation across discourse communities and time periods.

2. Methods

2.1. Corpus Linguistics and Genre Analysis

Our interest for the specialized knowledge produced by the IPCC emerged from our reading on climate change negotiations, in which it is quoted as an essential knowledge-provider, resulting in our hypothesis that this representation of the IPCC may be considered context-specific. A comparison of discourses from different types of specialized texts may indeed show that there is little consensus on the trustworthiness of this institution and that its results are considered with more distance in some contexts. Our aim is to better understand how the knowledge produced is recontextualized. According to Fairclough (2003, p. 51), recontextualization is "a movement from one context to another, entailing particular transformations consequent upon how the material that is moved, recontextualized, figures within that new context". When considering scientific knowledge, it is all the more interesting to analyse "contextualization procedures which make the communication of science not only possible, but useful and attractive" (Calsamiglia 2003, p. 146). It is particularly important to gain an understanding of these questions concerning climate change knowledge because of the "particularly multivoiced and multifaceted nature of the climate change debate, with a wide range of actors and voices, which causes multiple communication challenges" (Flottum 2017, p. 1).

The tools we draw on to explore this question derive from corpus linguistics (CL) combined with specialized genre analysis. CL traditionally refers to a quantitative approach to language based on the identification of statistically-significant occurrences in large electronic general language datasets. These quantitative methods have been fruitfully combined with other more qualitative perspectives on language such as genre analysis (Handford 2010), specialized discourse (Koester 2010) or Critical Discourse Analysis (Baker 2006), amongst others. Combining statistical methods with methods which are more attentive to context presents a number of advantages as summarised by Handford (2010, p. 256):

> Corpora have much to say about language, but they can be lacking in contextual interpretability; genres are intrinsically contextual entities, but their linguistic features may be under-exposed.

Even in specialized domains, which are traditionally considered less prone to diachronic terminological variation, an attention to historical context is now becoming increasingly widespread (Dury 2013; Dury, Picton 2009). In the case of a relatively recent specialized field like that of climate change, this temporal dimension can be seen as particularly relevant, since the theory on the topic as well as the discourse that describes it tend to evolve rapidly.

Our analysis is based on the comparative study of three discursive genres (Partington, Marchi 2015) that represent different levels of dissemination regarding knowledge produced by the IPCC. A genre is defined as a class of communicative events sharing a set of communicative purposes, a rationale that establishes constraints in terms of content, positioning and form (Swales 1990, p. 46). Comparing different genres thus implies taking into account the specific features of each genre, as well as an understanding of the discourse community (Swales 1990, p. 21) that produces it and its level of specialization. As for specialized discourse, the following definition by Gotti (2003, p. 24) is most appropriate to our purposes here:

> the specialist use of language in contexts which are typical of a specialized community stretching across the academic, the professional, the technical and the occupational areas of knowledge and practice.

Our first two corpora can undoubtedly be considered as being specialized in the field of climate change since they are produced by discourse communities familiar with this kind of knowledge, that is, the people in charge of communication during the conference of the parties (ENB corpus) and members of UN organizations (UN corpus). The third corpus, in contrast, is composed of general-interest press articles and represents popularization discourse (Gotti 2004). The three corpora are more precisely described in the following section.

2.2. Corpus constitution and description

2.2.1 Corpus constitution

As explained in paragraph 2.1, our analysis is based on the comparison of three different corpora composed of texts with distinct generic features. The questions of sample size and representativeness are central to the process of corpus building. Although different theories exist as to the ideal size of a corpus (Nelson 2010, p. 58), there seems to be a degree of consensus regarding the view that specialized corpora may be smaller than general corpora without losing any of their interest as an object of study, as explained by Koester (2010, p. 67):

> Where very large corpora, through their de-contextualisation, give insights into lexico-grammatical patterns in the language as a whole, smaller specialised corpora give insights into patterns of language use in particular settings.

Flowerdew (2004) lists six parameters according to which a corpus can be considered specialized: the specific purpose for compilation, the contextualisation in the case of a specific setting or communicative purpose, the genre, the type of text, the subject matter and the variety of English.

Our corpora may be considered specialized corpora according to three of the parameters listed by Flowerdew. Subject matter is the most obvious. The texts from our three corpora all focus on climate change. Genre is the second feature as our three corpora are genre based, as explained in paragraph 2.1. And, finally, contextualisation, since a specific setting and dates were among the elements which defined the constitution of our corpora. We gathered reports from international organisations and articles from the press, both in the United Kingdom and the United States. Concerning the dates, given that the highly mediatised events of COP15[1] and COP21[2] were milestones in climate change negotiations, we gathered documents published in the wake and aftermath of the event for each corpus, thus narrowing down our corpora to time frames situated between 2009-2010 and 2014-2017.

[1] Conference of the Parties 15, which took place in 2009 in Copenhagen.
[2] Conference of the Parties 21, which took place in 2015 in Paris.

Different levels of specialization may be observed in the constituted corpora. The first level belongs to a genre produced by a discourse community (Swales 1990) relatively familiar with climate change knowledge: people in charge of communication during the Conference Of the Parties. The documents included in these corpora are homogeneous and are all produced by the same institution: Earth Negotiation Bulletins (ENB) are summaries of the discussions which took place during COP negotiations. They are published by the International Institute for Sustainable Development and can all be downloaded directly from their website. All the documents published during the two time frames under scrutiny were selected for our two ENB corpora.

The second level of specialization concerns that of UN reports, which remain quite technical in their approach to the issues discussed, though they target a wider public. After identifying the UN institutions likely to publish such reports, we visited their websites and used their search engine to find documents of interest, which we classified according to the date of publication. These reports all deal with issues directly related to climate change and contain either the term "climate change" in their title or a related term like "climate mitigation", "climate deal", "climate action". Their length varies greatly, ranging from three- to two-hundred pages.

The third level of specialization concerns the dissemination and media coverage of the IPCC in the general-interest press, i. e. conveying specialized information to lay audiences (Moirand 2004, p. 84; Merhy 2010, p. 30). This may be viewed as a popularization corpus, as it implies recontextualization of the specialized information, as explained by Gotti (2014, p. 23):

> According to this new approach, popularization is thus not just seen as a category of texts, but as a recontextualization process that implies relevant changes in the roles taken on by the actors and institutions involved, and their degree of authoritativeness.

Recontextualization in the general press implies an adaptation of the discourse to different types of constraints such as "public interest and concern, market demands, the newspaper's ideological slant, and competition from other types of media" (Gotti 2014, p. 27). In this respect, it is worth pointing out that the treatment of science by the press is almost always marked by ideology, especially as pertains to the geographical area the articles are published in, the editorial line of the newspaper, or simply the dominant world vision at a certain time and place (Carvalho 2007; Boykoff, Boykoff 2007). It is thus primordial to take these parameters into account when constituting a press corpus. For the 2014-2017 sub-corpus, we chose four newspapers which represented two distinct geographical areas and diverse political opinions, as summarized in table 1:

Newspaper	Country	Political opinion
The Guardian	UK	left, liberal
The Daily Telegraph	UK	conservative
The New York Times	USA	liberal, democrat
USA Today	USA	centrist

Table 1
Description of the 2014-2017 corpus.

The 2009-2010 sub-corpus is smaller due to the unavailability of sources during this period; in fact, only the *Daily Telegraph* and *The New York Times* could be accessed for articles dating back to that time. Nevertheless, it may be considered a representative sample of a comparison corpus since it involves two different countries and divergent political leanings. The corpus was constituted with the help of the Europress database, using an extensive set of keywords including "climate change" and "IPCC", but also combinations of words such as "justice", "energy", "ecology" or "distribution". Our aim was to obtain a corpus representative of what was published on the topic of climate at the chosen periods.

2.2.2. General description of the corpus

In table 2 the data concerning our corpora are detailed.

	Number of words	Number of Documents	Publishing Organisations	Time Span
Earth Negotiation Bulletins	227,831	24	International Institute for Sustainable Development Reporting Services	2009-2010
	247,709	23		2015-2016
United Nation Reports	745,181	14	World Bank, UNEP, UNDP, REDD+, CBD[3]	2008-2010
	725,751	19		2014-2016
The Press	195,651	185	*The Daily Telegraph*, *The New York Times*	2009-2010
	423,459	429	*The Guardian*, *The Daily Telegraph*, *The New York Times*, *USA Today*	2014-2017

Table 2
General corpus description.

[3] Convention on Biological Diversity.

As seen in table 2, there is some discrepancy in sizes between our different sub-corpora. However, if the issue of size is central in deciding whether a corpus is representative of a specialized discourse, or not, the total number of words alone is not a sufficient criterion. In the BNC, 40,000 words were considered a target sample for different varieties of English. However, in the case of long reports, which is the case of our UN corpus, 40,000 words would potentially only cover one file. The number of files is also a relevant criterion and, for a genre in which files are longer, a greater number of words should be targeted. Although our first press sample is not vast in terms of words count, it nevertheless includes 185 articles, which represents a relatively high number of articles for the 2009-2010 period during which climate change coverage in the press was not as frequent as today. As Koester (2010, p. 71) affirms "many of the limitations of a small corpus can be counterbalanced by reference to the context" and the specificity of the texts gathered in this corpus counterbalances its relatively small size. As such, concerning the press corpus, we considered that the number of files and the representativeness of the sample of newspapers counterbalanced the relatively low word count. As for the ENB corpora, all the documents published at the dates of interest were included, thus eliminating the need for selection or extension of the corpora. In the case of the UN corpora, on the other hand, we did strive to find a balance between word count and number of documents.

2.3. Corpus Linguistics tools

Corpus linguistics offers a wide range of tools to the language analyst trying to describe distinctive features of a corpus. All of them are related to statistical considerations concerning lexical units and their relations. Our aim was to look at the context surrounding the phrase "International Panel on Climate Change" in its developed and abbreviated form, "IPCC". To do so, looking at collocates was a first step. "Collocates are useful in that they help to summarise the most significant relationships between words in a corpus" (Baker 2006, p. 118). Regarding the software TXM (Heiden *et al.* 2010) we used the log-likelihood feature. We needed to adapt the settings according to the number of words in our sub-corpora. The ratio named "indice" in the software, which indicates the statistical significance of the phenomenon observed, had to be adapted according to the size of the corpus. We used ratio 5 on the UN corpora where the word count is high. For the smaller press and ENB corpora, ratios 1 or 2 were more relevant, whereas for the 2009-2010 press corpus only ratio 1 gave results. We examined the contexts of the words in a window including ten words on the left and ten words on the right. We considered the first fifty lexical items after having deleted punctuation, numbers, but also verbs, as we wanted to deal with them separately. To sort verbs from the rest we used the part of speech tool integrated in the TXM software (Tree Tagger). We compared the results for each sub-corpus to highlight their respective specificities.

In a second stage we focused on collocate verbs. For each sub-corpus we considered the first ten results, including cases where the IPCC was in a subject or object position, to determine the type of action it is shown as being involved in. Verbs were then sorted according to semantic characteristics. Finally, going over all the concordance lines, we identified extracts where the IPCC words were quoted and determined the type of reported speech (Charaudeau 1992) used, the topic, the type of wording and the recontextualization process involved. In each case, the aim was to uncover patterns of meaning regarding the representation of the IPCC in the three genres over the two time periods.

3. Comparative analysis of collocates of IPCC in the three corpora

3.1. Concordance tables

Our preliminary step was to determine the frequencies of occurrences of "IPCC" in our corpora and how they were distributed across types of discourse and time periods.

	IPCC	Intergovernmental Panel on Climate Change
UN 2008-2010	367	106
UN 2014-2016	236	81
ENB 2009-2010	62	12
ENB 2015-2016	90	12
Press 2009-2010	42	60
Press 2014-2017	96	42

Table 3
Occurrences in the six sub-corpora.

Table 3 confirms the importance of the IPCC in the UN and ENB corpora. The acronym is clearly preferred to the developed form. In the press, however, the developed form is preferred in 2009-2010, while the acronym has more occurrences in 2014-2017. Acronyms are considered a sign of the degree of specialization of a discourse (Gotti 2003), which suggests that discourse on the IPCC in the press has grown to be more specialized over time. This is perhaps due to the institution becoming more familiar to the public. There are also marked differences in the way the institution is named in the different newspapers, as table 4 shows.

	IPCC	Intergovernmental Panel on climate change
The Guardian	59	22
The Daily Telegraph	77	29
USA Today	2	Ø
The New York Times	Ø	51

Table 4
Occurrences in the press corpus.

Although the two US newspapers tend not to use the acronym “IPCC”, several occurrences of the full name of the institution were found in the *New York Times* corpus. These appear most particularly in 2009-2010 (45 occurrences), – suggesting that the US press considers the public to be less familiar with this institution. Furthermore, there is a frequent tendency to explain what the IPCC is in the American press corpus, such as “a UN climate panel” (5 occurrences in the *New York Times*), “the UN climate change panel” (2 occurrences) “the climate change panel” (5 occurrences), or “the world’s top climate science panel” (1 occurrence).[4] In contrast, the British press mentions the acronym “IPCC”, but rarely adds an explanation as to what the institution is, or even its full name. It may suggest that the IPCC is considered a widely known institution in the UK, while it is not the case in the US.

3.2. Semantic categories identified in nouns

So as to determine the type of context in which the IPCC appears in each of our corpora, right and left noun collocates were identified. Table 5 shows the general results, among which several semantic categories could be identified. The acronym form was used for this calculation. To facilitate comparability of results, information is given about the number of co-frequencies and the ratios in brackets. Uppercase and lowercase fonts are distinguished.

[4] Since these formulations are difficult to identify systematically as referring to the IPCC, they were not taken into account in the analyses of coocurrents.

	2009-2010	2014-2017
ENB	Intergovernmental (11; 21), Panel (11; 21), Assessment (8; 14), Guidelines (5; 11), Inventories (5; 11), Pachauri (5; 10), Gas (5; 9), Report (7; 9), Greenhouse (5; 8), Expert (6; 8), phone (6; 8), Change (11; 7), Fourth (5; 7), Meeting (4; 7), AR4 (4; 7), Emirates (4; 7), Climate (13; 7), Special (6; 6), Arab (4; 6), Task (3; 6), Guidelines (3; 6), Abu (3; 6), Dhabi (3; 6), National (5; 6), location (5; 6), Report (4; 5), Rajendra (3; 5), 33rd (3; 5), approval (3; 5)	Stocktake (15; 15), Intergovernmental (9; 14), Hoesung (6; 12), Lee (6; 12), Panel (8; 12), assessments (7; 10), inputs (9; 10), global (16; 10), advice (6; 9), special (8; 9), assessment (9; 8), location (7; 7), Inventories (5; 7), Guidelines (5; 7), phone (6; 7), Change (11; 6), fax (6; 6), science (6; 6), reports (8; 6), AR6 (3; 5), Secretariat (7; 5), products (4; 5), Climate (12; 5), Special (4; 4), Panel (3; 4),
UN	AR4 (61; 99), Intergovernmental (43; 56), Panel (46; 52), Change (75; 41) Report (42; 33), Climate (75; 32), Assessment (32; 28), Fourth (22; 24), Cambridge (26; 20), WG (64; 19), ppm (15; 16), SPM (8; 14), Policymakers (7; 10), WMO (6; 10), According (12; 9), Synthesis (7; 9), Summary (9; 9), report (16; 9), evidence (16; 9), CO2-eq (9; 9), NY (8; 8), Press (12; 7), Geneva (7; 7), University (13; 7), Source (14; 6), Meteorological (5; 6), scenarios (10; 6), sensitivity (4; 5)	AR5 (42; 79), Intergovernmental (25; 33), Panel (51; 33), Change (43; 28), scenarios (37; 28), WGIII (12; 26), Climate (50; 24), Report (25; 19), report (25; 15), carbon (28; 14), database (13; 14), Assessment (16; 12), dioxide (16; 11), Synthesis (12; 11), warming (14; 11), Working (12; 9), Policymakers (5; 9), latest (6; 8), Summary (6 ; 8), Group (11 ; 8), Appendix (4; 7), Meyer (4; 7), Chapter (9; 7), scenario (14; 7), subset (5; 6), WG (3; 6), Cambridge (8 ; 5), Geneva (5; 5), GWPs (5; 5), Switzerland (5; 5), scientific (6; 5), chance (6; 5)
Press	chairman (4; 4), claim (2; 2), climate (6; 2), credibility (2; 2), head (3; 3), inquiry (2; 2), management (3; 4), panel (5, 6), report (6; 3), review (3; 3), errors (2; 2), organisations (2; 3), change (6, 4), IAC (2; 4), Intergovernmental (6; 6), Pachauri (5; 4)	balance (2; 2), chair (2; 2), chairman (4; 3), Climate (20; 12), crisis (4; 2), department (2; 2), draft (3; 2), group (4; 2), member (3; 2), panel (3; 2), Panel (8; 13), report (38; 37), route (2; 2), science (5; 2), heatwaves (2; 2), reports (3; 2), shortages (3; 4), thanks (2; 2), Change (2; 22), College (2; 2), Imperial (2; 3), Intergovernmental (20; 34), Jean (2; 4), Jouzel (2; 4), Masson-Delmotte (2; 4), Pachauri (3; 4)

Table 5
List of "IPCC" collocates.

Two main semantic categories emerge in the three corpora, namely work and results. The list of collocates of IPCC in the two ENB corpora share many lexical units in common: Intergovernmental, Panel, Assessment, Guidelines, Change, Special. All of these figure in titles published by the IPCC which are quoted in the ENB, like, for example the *Guidelines for National Greenhouse Gas Inventories*, quoted in both, or Special Reports on different topics treated by the IPCC. In the first ENB corpus the *Special Report on Renewable Energy Sources* is quoted. In the second, a task force named Task Group on Data and Scenario Support for Impact and Climate Analysis is mentioned. Sometimes, the reports published by the IPCC are referred to in more general descriptive terms like in the expression "IPCC assessment", "IPCC report", which explains the presence of these lexical units in both columns of the table. On the same line "IPCC inputs" and "IPCC products" figure in the second corpus. An unexpected lexical unit found in both columns is “phone”. This is because, at the end of the ENB documents, information concerning the next IPCC meeting indicating location, address, fax number and phone number is provided. This also explains the presence of lexical units like “fax”, “location” and names of countries and cities.

If we focus on the differences between the first and the second ENB corpora, we find different proper names like Rajendra Pachauri in the first and Lee Hoesung in the second. This is because they are the chairmen of the IPCC, Rajendra Pachauri from 2009-2010 and Lee Hoesung from 2014-2017. Other lexical units that appear in the second corpus are linked to a change in the international context surrounding the IPCC. The presence of “Stocktake” and “Global” is the most obvious. “Global Stocktake” was a term coined in the article 14 of the Paris Agreement to refer to a key mechanism of the Agreement. It implies that all signatory parties assess the collective progress towards achieving the purpose of the agreement and its long-term goals every five years. It is discussed at length during the COP22 as the ENB reports published on this conference testify, one of the key points being to work out how the IPCC can help with this Global Stocktake. Another collocate which appears in the second corpus is “Science” notably in extracts in which people speaking for the IPCC intervene to highlight how important it is to base decisions on science.

The UN collocates are comparable to those found in the ENB. Indeed, most of them are lexical units extracted from titles published by the IPCC. We also find detailed references about the publications that explain the presence of lexical units such as “Geneva”, “Press”, “Cambridge”. Parts of reports are also referred to with the acronym “SPM”, standing for Summary for Policy Makers. The acronym WG is also used to allude to a part of the report, as it stands for Working Group referring to the three IPCC working groups, each with its specific disciplinary specializations and each publishing a separate part of the assessment report. “Evidence” and “scientific” refer to the work of the IPCC, highlighting its scientific basis. “Chance” appears in the expressions “likely chance” and “66% chance”, used to estimate how likely it is for an event planned by the IPCC to take place. However, a notable difference with the ENB is that we find more collocates concerning precise results published by the IPCC. “Ppm” and “CO2-eq”, for instance, are found in extracts concerning estimates on maximum concentrations. “Carbon” and “dioxide” also appear in statements about calculations on maximum carbon concentrations provided by the IPCC. “IPCC scenarios”, “scenario databases”, “subset of scenarios”, “limiting warming to 2°” are other examples of references to results found in the IPCC reports. The comparison of collocates in the UN and the ENB corpora tends to indicate that in both, the IPCC is a reference for specialized knowledge on climate change.

In the press corpora, the semantic category of work is mainly mentioned through the documents produced by the IPCC, especially with the words “report” or “draft”. The lexical unit “report” is largely predominant as a coocurrent in 2014-2017. References to the participants via their status, such as “chairman”, “head” or “member”, are also present in both corpora, likewise via the proper nouns, like, for instance Rajendra Pachauri and Jean Jouzel, the chair and vice-chair of the IPCC.

The semantic category of results is also treated differently. While the ENB corpus refers to the results of the IPCC mainly to insist on scientific validation, scientific processes are never discussed in any of the corpus newspapers. In 2009-2010, there is no mention of the scientific results obtained by the IPCC, except through the titles of the documents produced by the institution. In 2014-2017, the focus lies on the topics dealt with in the reports, with coocurrents such as “balance”, “crisis”, “food”, “science” or “heatwave”, for instance:

> In a report to be released today, the IPCC warns of flooding, droughts, heatwaves and food shortages that are likely to result from rising temperatures and extreme weather patterns. (Telegraph 2014).[5]

Interestingly, while results are often presented as facts, the negatively connoted coocurrent “claim” used by *The Daily Telegraph* in the 2009-10 corpus seems to reflect doubts as to the reliability of the specialized knowledge of the IPCC itself:

> The IPCC claimed that up to 40 per cent of the Amazonian forests could be badly affected by global warming. The claim was tracked back to a report by WWF and the IPCC was criticised for using environmental groups as sources. The claim continues to be contested. (Telegraph 2010)

The single occurrence of this critical evaluative stance with regard to the IPCC’s specialized knowledge may be accounted for by the conservative political tendency of *The Daily Telegraph*. The fact that it is absent from the more recent corpus may also indicate that representations of the IPCC have evolved, as section 4.3 will develop.

In the press corpus, there is a clear diachronic evolution between the sub-corpora, the 2009-2010 corpus being the only one to contain references

[5] Quotes are identified with the name of the author organisation or newspaper and the year of publication in the whole paper.

to the lack of credibility of the IPCC, with coocurrents such as “credibility”, “inquiry”, “review” or “errors”, for instance:

> Errors in the 2007 assessment report, including a prediction that the Himalayan glaciers could disappear by 2035, have threatened to overshadow the United Nations' message that climate change is a significant threat requiring urgent collective action. (NYT 2010)

These references are to be interpreted with regard to the Climategate scandal: in 2009, internal messages were hacked from the Climatic Research Unit of the University of East Anglia which worked closely with the IPCC. Although the emails mainly concerned everyday communication between scientists, certain phrases were interpreted as proof that evidence of anthropogenic climate change had been manipulated and subsequently used by climate deniers to defend the idea of a scientific conspiracy on climate change. The fact that the scandal broke only weeks before the Copenhagen summit, which gave a large coverage to climate discussions, may explain why it lastingly damaged the reputation of the IPCC. Though references to controversies on climate in the 2009-2010 corpus are unsurprising, this type of coocurrent completely disappeared in 2014-2017.

To conclude on this point, while the same semantic categories appear in all the three corpora, their content differs. Though in each case collocates suggest that the IPCC is mainly referred to in terms of its work and the results of its publications, the results are described in greater detail in the UN corpora. The ENB corpora provide fewer technical details. The press tends to focus on some general aspects of the publications and it does not provide any technical details or explanations of the scientific processes involved. Finally, the diachronic variation tends to suggest that the position of the IPCC as an essential knowledge-provider on climate change has been strengthened between the first time period and the second.

3.3. Semantic categories identified in verbs

We used the part of speech identification tool TreeTagger, which is integrated to TXM, to highlight the verbs co-occurring with “IPCC” in the right and left contexts (3/3), retaining the first ten results for each corpus. Table 6 summarizes the results, with co-frequencies and ratios indicated between brackets:

	2009-2010	2014-2017
ENB	Use (4; 4), **observed** (2; 4), follow (2; 2), **discussed** (4; 2), **reported** (4; 2)	**Inform** (9; 11), substantiated (2; 4), derived (2; 4), **Inform** (2; 3), provide (7; 3), **determined** (2; 3), refine (2; 3), distinguishing (2; 3), **relayed** (2; 3), **questioned** (2; 2)
UN	**Reported** (7; 8), **estimated** (6; 7), **published** (7; 6), **based** (6; 5), **concluded** (4; 4), **stated** (3; 4), **evaluate** (3; 3), **reported** (4; 3), **described** (3; 3), **underlined** (2; 3)	**Based** (18; 8), **indicated** (4; 6), **concludes** (4; 5), **published** (4; 4), limiting (4; 3), **states** (3; 3), **shown** (5; 3), **analyses** (2; 3), **estimated** (3; 2), **concluded** (2; 2)
Press	was (4; 1), made (2; 2), **said** (5; 1), does (2; 1)	**write** (2; 3), **concluded** (2; 3), **found** (3; 3), **warned** (2; 2), **finds** (3; 6), **says** (6; 4), **warns** (5; 8)

Table 6
Verb collocates of “IPCC”.

Most verbs in the three corpora are reporting verbs, which figure in bold in our table. All in all, we found a strong diversity of reporting verbs. According to Hyland (2002), reporting verbs can be categorized according to process types and evaluative stances. Roughly speaking, process types refer either to research acts, cognition acts or discourse acts. Evaluative functions can either reveal supportive stances towards the author’s claims (“affirm”, “explain”, “note”), tentative stances (“postulate”, “hypothesize”), critical stances (“evade”, “exaggerate”) or neutral stances (“describe”, “say”).

In the ENB corpus, the reporting verbs refer either to research acts (“observe”, “determine”) or to discourse acts (“discuss”, “report”, “inform”, “relay”, “question”). The stance is either neutral or supportive. There are no examples of tentative or critical stances. Two exceptions to reporting verbs are worth noting: “to use” and “to distinguish”. By checking the concordance lines, we observed that the verb “to use” appeared four times in the following type of extract: “On common metrics, parties discussed whether to use the Intergovernmental Panel on Climate Change’s (IPCC) Fourth Assessment Report (AR4) for GWPs” (ENB2016). In one case, it is the IPCC that is asked to use another document as a reference: “Parties discussed requesting the IPCC to use scientific developments on wetlands to fill gaps in the 2006 IPCC Guidelines” (ENB2016). This suggests that the negotiators also consider intervening in the work of the IPCC and issuing guidelines. In the case of “distinguishing”, which is not a reporting verb, the concordance lines revealed the verb was used with the verb “called for” in the context of advice given on how to use IPCC results:

> On sources of input, many agreed that the scientific inputs should be mainly derived from the IPCC and called for distinguishing between sources, such as the IPCC, and information. (ENB 2017)

Overall, in the ENB corpus, some occurrences suggest that the role of the IPCC is not just to inform but also to be a service provider (suggested with the verb "to use") and a collaborator.

It seems noteworthy that in the case of the UN, we find nearly exclusively reporting verbs, with one exception in the second. Like in the ENB corpora the verbs refer to research acts ("estimate", "base", "conclude", "evaluate", "analyse") and to discourse acts ("report", "publish", "state", "describe", "underline", "indicate", "show") and the evaluative function is mainly supportive or neutral. The verb "estimate" may be seen as having a tentative stance although this is context-dependent. The verb "show" expresses a strong supportive stance.

In the press, quotes are used by journalists to delegate the responsibility of the content to their sources so as to appear more neutral, but the authors may nevertheless "influence our interpretation by their use of adverbs, adjectives and introductory verbs" (Komur 2004, p. 61). The analysis of these verbs is thus crucial to understand how a journalist represents the source it evokes. The small size of the 2009-2010 press corpus makes it difficult to obtain results on such a specific search. However, in the 2014-2017 corpus, reporting verbs are the most frequent, belonging to the category of research acts ("find", "conclude") or discourse acts ("say", "write", "warn") (Hyland 2002). Most of the reporting verbs used are neutral, which do not suggest any specific interpretation of the quote to readers, as it is the case of "say", "write" or "find". As research act verbs, "conclude" or "find" clearly highlight the scientific nature of the IPCC's work and findings, for example:

> The IPCC concluded in 2013 that even if the increasingly quixotic-looking safe limit of 2C of global warming were somehow achieved by the Paris talks, the sea would continue to wash over Kiribati and the Marshall Islands. (Guardian 2015)

These two verbs are used about twice as much in the *Guardian* and the *New York Times* as in the other two newspapers. This could be because they present a more positive view of the IPCC's work and endorse its scientific nature.

A similar discrepancy between the newspapers appears with the verb "warn", which belongs to the category of supportive reporting verbs. This verb underlines the role of the IPCC as not only a provider of knowledge, but also a stakeholder seeking to convince people of the dangers of climate change. The fact that it is more frequently used in left-leaning newspapers is thus consistent with their environmental commitment.

The collocate verbs used with the IPCC confirm the idea that it is often a reference since its results are quoted in all three corpora. However, we find more examples of a supportive stance in the ENB and UN corpus than in the press corpus, in which the majority of reporting verbs are neutral. Besides, verbs in the ENB corpus reflect the fact that the IPCC is also a collaborator whose work is used and discussed. To better understand these various roles, it is important to take a closer look at citations present in our different corpora.

4. Themes and aims of IPCC's quotes

4.1. Quotes in the ENB corpus

In the first time period sub-corpus we found eight quotes from the IPCC, all using indirect reported speech. In three cases, it was the IPCC chair who was quoted, as in the following extract:

> Rajendra Pachauri, Chair of the Intergovernmental Panel on Climate Change (IPCC), highlighted the consequences of failure to implement climate change mitigation policies on the basis of the Fourth Assessment Report. (ENB2015)

In the five other cases the IPCC was mentioned as an organisation, with no indication as to who was speaking. Concerning the themes found in the quotes, in three cases, the aim of the quote was to highlight the urgency of the situation and the necessity for policy makers to act quickly to limit climate change, as in the following example:

> According to the Intergovernmental Panel on Climate Change (IPCC), the effects of climate change are already being observed and scientific findings indicate that prompt action is necessary. (ENB2015)

In two cases, the IPCC was quoted as giving details about how deforestation increases carbon emissions. In one case the theme was the vulnerability of developing countries and the necessity of helping them. In two cases, details were given about the work of a special Task Group.

In the second time period, only three quotations were retrieved, two by the new IPCC Chair, Hoesung Lee, and one by his co-chair Hans-Otto Portner. As in the previous time period, the IPCC is mainly quoted to highlight that science is clear on the extent of climate change and the urgency to act. The following example seems significant from this point of view:

> Hans-Otto Pörtner, Co-Chair, IPCC Working Group II, stated that the role of science is to reduce uncertainties and lamented that thus far there has been a degree of societal inertia and inaction to address issues highlighted by science. (ENB2017)

Overall, the IPCC is the voice of science in the ENB corpus. Quotations from the IPCC are used to confirm the reality of the phenomenon of climate change. The validity of their statements is presented as unquestionable. Even in the sub-corpus from 2009-2010 where the IPCC was subject to criticism because of the Climategate, the quotation is used to express a clear endorsement of the IPCC's results:

> Pointing to a recent incident involving the theft of emails from scientists at the University of East Anglia in the UK, Pachauri highlighted the IPCC's record of transparent and objective assessment. (ENB 2015)

During COP negotiations, IPCC reports are seen as presenting unambiguous scientific results which come as a warning to political representatives about the urgency to agree on solutions and implement them to fight the related risks.

4.2. Quotes in the UN corpus

In the 2009-2010 UN corpus, there are 110 occurrences of "IPCC" in brackets. In these cases, it is used as a reference to back up a statement. In the 2014-2017 corpus there are 57. This confirms that the IPCC is an essential source of knowledge in the UN reports. By looking through concordance lines and using punctuation marks and reporting verbs, we retrieved all occurrences of explicit reported speech whether it be in the direct or indirect form. For each statement we identified the main theme. Table 7 shows the most and least present themes in the quotations.

	UN1 Direct	UN2 Direct	UN1 Indirect	UN2 Indirect	Total
Scenario Results		2	3	10	15
Specific Ecosystems Impacted			10	1	11
Mitigation	2		3	5	10
Climate Change General	2	1	3	2	8
Agriculture and Food Security			2	3	5
Adaptation	2	1			3
Vulnerability	1	2			3
Small Islands	2				2
Human Impact on Climate Change	1		1		2
Carbon Budget		1		1	2
Role of IPCC	1				1
Methods for Assessment			1		1
Absorption			1		1
Economic Loss		1			1
Developing Countries				1	1
Cities' Impact				1	1
TOTAL	11	8	24	24	67

Table 7
Themes in UN quotations.

To speak about the themes in table 7 the IPCC is explicitly considered a reference. The importance of the IPCC scenario results is clear as it is the main topic of the quotations. Here is an example of indirect quotation presenting scenario results: “IPCC (2007e) estimates that carbon prices of US $ 20-80 / tonnes CO2-eq would be required by 2030 to aim at achieving stabilization at around 500 ppm CO2-eq by 2100”. The IPCC is also regularly quoted as an authority on the specific impacts of climate change on an ecosystem, as in the following example:

> According to the IPCC AR4, both tropical and temperate grasslands are sensitive to variability and changes in climate, which are likely to have strong effects on the balance between different life forms and functional types in these systems. (CBD 2009)

There is a higher frequency of indirect quotations than direct quotations. It is interesting to note that certain themes appear exclusively in direct quotations, such as, for example, those related to adaptation and vulnerability. This is probably because these are concepts coined by the IPCC and passages devoted to defining them are therefore quoted in direct speech, as in the following example:

> The Intergovernmental Panel on Climate Change IPCC defines adaptation as “adjustments in ecological, social, or economic systems in response to actual or expected climatic stimuli and their effects or impacts. This term refers to changes in processes, practices, and structures to moderate potential damages to or benefit from opportunities associated with climate change”. (UNDP 2009)

We noticed seven other occurrences in which direct quotes are used to present an IPCC definition, the exact wording of definitions being essential. This probably explains the high number of definitions among direct quotations. It may also be explained by a willingness on the part of UN institutions to establish terminological coherence by using similar terms and defining them identically. Another case in which direct quotes seem to be preferred to indirect ones is when introducing a controversial question. Small island states are seen as being particularly at risk as they are among the first directly impacted by climate change. The question of the status of their displaced inhabitants is therefore difficult to solve, as made explicit in the following extract:

> As the IPCC identifies, “the costs of overall infrastructure and settlement protection are a significant proportion of GDP and well beyond the financial means of most small island states”. (UNEP 2010)

When evoking controversial questions, using the exact wording of the organisation quoted may facilitate a faithful presentation of its ideas. This is essential in the context of a debate. Other themes seem to be exclusively presented in indirect reported speech, such as, for example, impacted ecosystems. These questions may be seen as less controversial, which would explain why the exact wording used by the IPCC is not crucial.

Overall, looking through quotes from the IPCC in the UN corpus, there is no evidence of a questioning of the IPCC as an essential knowledge-provider. On the contrary, the great value of its findings is asserted several times, as in the following example:

> The Intergovernmental Panel on Climate Change (IPCC), which was set up by the World Meteorological Organization and the United Nations Environment Programme, is widely recognized as the principal authority for objective information on climate change, its potential impacts, and possible responses to these. (WTO 2009)

In view of these elements, we may conclude that its reports are presented as landmarks for the development of knowledge on climate change.

4.3. Quotes in the press corpus

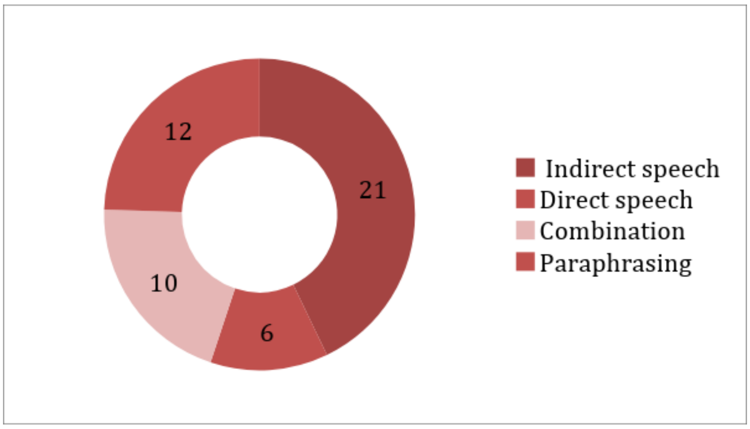
Paragraph 4.2 has shown that most of the reporting verbs found in the press corpus were neutral. However, the journalist's attitude towards a quote is also reflected in the types of quotes used (Peynaud 2011), so that direct speech, indirect speech or paraphrasing are found in specific contexts. Indeed, quotation marks show the distance journalists take with the quote in a process of legitimization of the discourse (Adam, Lugrin 2006). Whether it is a question of quotes expressing opinion or highly specialized ones, using quotation marks allows journalists to mask their own voice behind those of more legitimate speakers (Marnette 2004). For instance, the wording of quotes is generally characterized by a low degree of technicality, accessibility to the general public being one of the aims of the general-interest press. The few technical terms that are used are mainly found in direct speech, which legitimizes the use of a term by clearly attributing it to an expert source as in the following example.

> The IPCC warns: "A large fraction of both terrestrial and freshwater species faces increased extinction risk under projected climate change during and beyond the 21st century, especially as climate change interacts with other stressors, such as habitat modification, over exploitation, pollution, and invasive species." (Telegraph 2014)

The low number of quotes related to the IPCC in the corpus made it possible to classify them into four main categories: direct, indirect, combinations of both or paraphrases. Analysing the content and the aim of the quotes with regard to the types of quotes gives a more specific insight into how the knowledge produced by the IPCC is disseminated in the press. In addition to direct and indirect speech, combinations of both types of reported speech are often found in the corpus; in these cases, a short segment is between quotation marks in the middle of a sentence in indirect mode, for instance:

> The more the climate warms, the more people will experience "water scarcity", the IPCC finds. If carbon emissions remain high, droughts will become more frequent in dry regions by the end of the century although "water resources are projected to increase at high latitudes". (Telegraph 2014)

Occurrences of paraphrasing were also found, where the author summarizes the main ideas of the report, for example: "But the IPCC scientists call for trebling of the use of renewables takes on a new force if the world's fastest-growing polluter has made such a big move." (Guardian 2014) The results of the classification into four categories are shown in graph 1.



Graph 1
Types of quotes in the press corpus

The IPCC is sometimes the direct source of the quotes, particularly with regard to its prominent members. Rajendra Pachauri, chair of the IPCC from 2002 to 2015, is mentioned 77 times in the corpus. Jean Jouzel, vice-chair of the IPCC, is mentioned 3 times, and Valérie Masson-Delmotte, vice-chair of work-group 1, is mentioned twice as a collocate of "IPCC". Press quotes mainly deal with two topics: the dangers of climate change and the possible solutions, in equal proportions. Uncertainty about IPCC results is never explicitly mentioned. This is consistent with news values, as Carvalho (2007, p. 229) points out: "Uncertainty is a difficult issue for reporters, as news values of clearness and unambiguity demand 'facts' and lead to a streamlined image of scientific knowledge". However, although uncertainty is never the main topic of the quotes, it may appear in a more implicit form, through modals, for instance:

> For example, the 2013 Intergovernmental Panel on Climate Change (IPCC) report says we can expect a warming of between 1.5° C and 4.5° C if we double atmospheric CO2 levels, but also acknowledges that the longer term warming (over centuries to millennia) "could be significantly higher" than that. (Guardian 2016)

A combination of direct and indirect speech is used here, with facts expressed in indirect speech and the modal phrase between quotation marks. This combination illustrates the fact that the journalist does not take responsibility for the hypothesis mentioned but attributes it to the IPCC directly in direct mode.

In contrast, several quotes, especially in the 2014-2017 corpus, stress the fact that the latest IPCC reports show more certainty about the possible consequences of climate change, for example:

> The latest IPCC report, the first update in seven years, confirms that climate scientists appear *more certain than ever before* that human behaviour is the key culprit for global warming.[6] (Telegraph 2014)

Results in this regard are remarkably similar between newspapers and periods, perhaps reflecting the fact that this treatment of uncertainty is mainly due to the genre of press articles.

The main aims of the quotes are first, to inform the public about the results obtained by the IPCC in a neutral manner by reporting facts and figures. 41% of the quotes correspond to such facts, while 43% refer to warnings expressed by the IPCC on the dangers of climate change. Finally, some quotes expressing criticism of the IPCC may take different forms. Facts and figures, which are non-polemical information, are often quoted in indirect mode or with a paraphrase, for instance:

> In a report to be released today, the IPCC warns of flooding, droughts, heatwaves and food shortages that are likely to result from rising temperatures and extreme weather patterns. (Telegraph 2014)

In contrast, direct quotes tend to contain an evaluation of the situation, especially in the case of warnings, since journalists are meant to preserve neutrality and thus cannot explicitly give an evaluation of the situation in their own voice. This process appears particularly clearly in combinations as, for instance:

[6] Emphasis added.

> A report by the Intergovernmental Panel on Climate Change (IPCC) found it was still just possible to limit the rise in average global temperatures to 3. 6F (2C) by 2100, the level beyond which experts say the effects will be "dangerous". (Telegraph2014)

However, there are clear differences in the way the types of quotes are used by different newspapers. Indeed, the *Guardian* and the *New York Times* use more paraphrasing and indirect speech than the *Daily Telegraph* and *USA Today*. This division is consistent with their political leanings and clearly shows the attitude of the journalists towards the content of the IPCC. In the first two newspapers, facts and figures are often quoted in indirect speech modes or paraphrasing them. Only about 20% of the quotes include quotation marks, showing that the journalist integrates the IPCC's words into his own discourse, for instance:

> Depending on how sharply the world cuts carbon emissions, the Intergovernmental Panel on Climate Change (IPCC) predicts the global mean sea level will rise by 26-82cm between now and 2100. (Guardian 2015)

The IPCC is here considered a trustworthy source of information on climate change and its scientific nature is highlighted by the use of figures. In the *Daily Telegraph* and *USA Today*, in contrast, 50% of quotes contain quotation marks, thus underlining the fact that the journalist shows some distance with regard to the IPCC's words as, for instance: "'All aspects of food security are potentially affected by climate change, including food access, utilisation, and price stability,' the IPCC says." While left-leaning newspapers quote the knowledge produced by the IPCC as facts, the other newspapers chose to show more distance and to clearly attribute the words to the institution. This phenomenon is consistent throughout the two periods.

Finally, some quotes reflect a certain criticism of the IPCC's work, essentially in paraphrasing mode:

> If the IPCC report admits that the environmental crisis will not be felt equally across the population, it didn't push the argument until its logical political conclusion. (Guardian 2014)

However, there are only three quotes in the whole corpus reflecting such opinions, disseminated in three different newspapers, which seems to indicate that on the whole the press expresses little direct criticism of the IPCC. The press does, nevertheless, reflect the state of the opinion at a certain period of time. In the 2009-2010 corpora, the IPCC's credibility was questioned because of the "Climategate" and an exaggerated claim, according to many studies, that glaciers would disappear by 2035. This general loss of confidence in the IPCC is reflected in the immediate context surrounding the acronym and the full name of the institution:

> Growing public skepticism has, in recent months, been attributed to news reports about e-mail messages hacked from the computer system at the University of East Anglia in Britain (characterized as showing climate scientists colluding to silence unconvinced colleagues) and by the discoveries of alleged flaws in reports by the Intergovernmental Panel on Climate Change. (NYT 2010)

The "growing public scepticism" mentioned here highlights the wide-scale consequences the scandal had on the image of the IPCC with the general public. In 2014-2017, scepticism towards IPCC results is presented as a minority opinion as the following extract testifies:

> And finally, he shared his views on what he described as the Obama administration's lies about global warming. It doesn't concern Mr. Murray that his climate views run counter to *the scientific consensus, which has found that warming is "unequivocal", according to the latest report from the Intergovernmental Panel on Climate Change.* "This global warming hoax," Mr. Murray told the students, "it's a movement that is destroying America." (NYT 2016)

In this example, the journalist refers to the "scientific consensus" that global warming is a reality. The minority opinion of Mr. Murray, who defends the opposite opinion, is inscribed within the dominant journalistic discourse and is largely discredited. The legitimacy of the institution is underlined by phrases such as "we know" or "thanks to", adjectives relating to the report like "comprehensive", or references to the many scientists who worked on the report and the scientific consensus that surrounds it.

5. Conclusion

Overall, the IPCC is presented as the main reference on climate change knowledge in the six sub-corpora analysed. All three sources disseminate the knowledge produced by the IPCC and stress the legitimacy of this source. However, the generic specificities of the three corpora show a contrasting use of the IPCC as a knowledge provider. In the UN and ENB corpora, the IPCC is mainly quoted to back up arguments about the reality of climate change and the necessity to fight it. Details are given about results of IPCC reports and the scientific nature of its work is highlighted. In the press corpus, the recontextualization process implies the use of fewer scientific terms and quotations which concern less technical aspects of its work. It is also the corpus in which there are most of the references to differing opinions. The diachronic perspective highlights an evolution in the representation of the

IPCC. Indeed, while the 2009-2010 corpora refer to the controversies that undermined the legitimacy of the institution at the time, scepticism towards its results is shown as a minority opinion in 2014-2017.

The use of six sub-corpora, representing three levels of dissemination over two time periods enables a varied approach to IPCC knowledge. Comparisons between the corpora have shown varied representations of the IPCC, depending on the level of specialization and political leaning in the case of newspapers. Knowledge-providers do not have the same role in the press as in a specialized corpus like that of the UN reports. To build on these results, we believe it would be interesting to use the collocates identified in the immediate contexts of the word IPCC to see if other knowledge-providers can be identified in our corpora. Similar methods could then be used for the three corpora, namely the analysis of collocates and quotations, to define the role of different stakeholders involved in climate negotiations and activism. Climate change is a field in which the production of knowledge is all the more important as it is an essential step towards the negotiation of an international agreement to limit climate change. A better understanding of knowledge dissemination in this field thus seems important. Although the IPCC figures as an essential source, it would be of interest to map the dissemination of other less widely known sources of knowledge so as to question or confirm the uncontested character of this leading institution.

Bionote: Camille Biros is an Associate Professor at the Health Faculties in Grenoble (Université Grenoble Alpes) where she teaches English for Medicine, Pharmacy and Biotechnology. Her research focuses on health and environmental discourse, terminology and communication strategies.
Caroline Peynaud is an Associate Professor in English at the Applied Modern Languages department in Grenoble (Université Grenoble Alpes), where she mainly teaches economic English and grammar. Her research focuses on journalism as a specialised domain, more precisely on the specialised culture of journalists in the United States, on the discourse they produce and on the genres and the terminology that characterise it.

Authors' addresses: camille.biros@univ-grenoble-alpes.fr; caroline.peynaud@univ-grenoble-alpes.fr

Acknowledgements: We would like to thank the CORLI (CORpus, Langues, Interactions) consortium and the Université Grenoble Alpes Data Institute for providing support to build the corpus. We would like to thank Shaeda Isani for her precious proofreading of the article.

References

Adam J.-M. and Lugrin G. 2006, *Effacement énonciatif et diffraction co-textuelle de la prise en charge des énoncés dans les hyperstructures journalistiques*, in "Semen. Revue de sémio-linguistique des textes et discours", 22. http://semen.revues.org/document4381.html (17.08.2009).

Aykut S. and Dahan A. 2015, *Gouverner le climat ? Vingt ans de négociations internationales*, Presses de Sciences Po, Paris.

Baker P. 2006, *Using Corpora in Discourse Analysis*, Bloomsbury Academic, London.

Boykoff M.T. and Boykoff J.M. 2007, *Climate Change and Journalistic Norms: A Case-Study of US Mass-Media Coverage*, in "Geoforum" 38 [6], pp. 1190-1204.

Calsamiglia H. 2003, *Popularization Discourse*, in "Discourse Studies" 5 [2], pp. 139-146.

Carvalho A. 2007, *Ideological cultures and media discourses on scientific knowledge: re-reading news on climate change*, in "Public understanding of science" 16 [2], pp. 223-243.

Charaudeau P. 1992, *Grammaire du Sens et de l'Expression*, Hachette, Paris.

Colson R., de Munagorri E., Denis B., Leclerc O., Rousseaux S. and Torre-Schaub M. 2009, *Expertise et Gouvernance du Changement Climatique*, LDGJ, Paris.

Dury P. 2013, *Que montre l'étude de la variation d'une terminologie dans le temps. Quelques pistes de réflexion appliquées au domaine medical*, in "Debate Terminológico" 9, pp. 2-10.

Dury P. and Picton A. 2009, *Terminologie et diachronie : vers une réconciliation théorique et méthodologique?*, in "Revue Française de Linguistique Appliquée", XIV [2], pp. 31-41.

Fairclough N. 2003, *Analysing discourse, textual analysis for social research*, Routledge, London, New York.

Flottum K. (ed.) 2017, *The Role of Language in the Climate Change Debate*, Routledge, New York.

Flowerdew L. 2004, *The Argument for Using English Specialised Corpora to Understand Academic and Professional Settings*, in Connor U. and Upton T. (eds), *Discourse in the Professions: Perspectives from Corpus Linguistics*, Benjamins, Amsterdam, pp 11-33.

Gotti M. 2003, *Specialized discourse: linguistic features and changing conventions*, Peter Lang, Bern.

Gotti M. 2004, *Specialised Discourse in Multilingual and Multicultural Contexts*, in "ASp" 45-46, pp. 5-20.

Gotti M. 2014, *Reformulation and Recontextualization in Popularization Discourse*, in "Ibérica" 27, pp. 15-34.

Handford M. 2010, *What Can a Corpus Tell Us About Specialised Genres*, in McCarthy M. and O'Keefe A. (eds.), *The Routledge Handbook of Corpus Linguistics*, Routledge, New York, pp. 255-269.

Heiden S., Magué J-P. and Pincemin B. 2010, *TXM: Une plateforme logicielle open-source pour la textométrie – conception et développement*, in Bolasco S., Chiari I. and Giuliano L. (eds.), *JADT 2010 – Statistical Analysis of Textual Data Proceedings of 10th International Conference*, LED Edizioni Universitarie, Roma, pp. 1021-1032.

Hyland K. 2002, *Activity and evaluation: reporting practices in academic writing*, in Flowerdew J. (ed), *Academic Discourse*, Longman, London, pp. 115-130.

Koester A. 2010, *Building Small Specialised Corpora*, in McCarthy M. and O'Keefe A.

(eds.), *The Routledge Handbook of Corpus Linguistics*, Routledge, New York, pp 66-79.

Komur G. 2004, *Les modes du discours rapporté dans la presse et leurs enjeux polyphoniques*, in “Pratiques” 123 [1], pp. 57-74.

Marnette S. 2004, *L’effacement énonciatif dans la presse contemporaine*, in “Langages” 156 [4], pp. 51-64.

Merhy L. 2010, *La vulgarisation dans les médias: sciences et émotions*, in “Communication, lettres et sciences du langage” 4 [1], pp. 29-41.

Moirand S. 2004, *De la médiation à la médiatisation des faits scientifiques et techniques: où en est l’analyse du discours?*, in “Colloque Sciences, Médias et Société” 15-17. http://sciences-medias.ens-lsh.fr/article.php3 ?id_article=59 (24.04.2013).

Nelson M. 2010, *Building a Written Corpus*, in McCarthy M. and A. O’Keefe (eds.), *The Routledge Handbook of Corpus Linguistics*, New York: Routledge, pp 53-65.

Partington A. and Marchi A. 2015, *Using Corpora in Discourse Analysis*, in Douglas B. and Reppen R. (eds.), *The Cambridge Handbook of English Corpus Linguistics*, Cambridge University Press, Cambridge, pp. 216-234.

Peynaud C. 2011, *Parole rapportée et positionnement discursive dans la presse américaine: analyse de l’utilisation des citations dans des commentaires politiques*, in “ASp” 59, 43-64.

Swales J. 1990, *Genre Analysis: English in Academic and Research Settings*, Cambridge University Press, Boston, pp. 21-32.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 205-224
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p205

TRANSLATING (IM)PERSONALISATION IN CORPORATE DISCOURSE
A corpus-based analysis of Corporate Social Responsibility reports in English and Italian

SARA CASTAGNOLI[1], ELENA MAGISTRO[2]
[1]UNIVERSITÀ DI MACERATA, [2]UNIVERSITÀ DI BOLOGNA

Abstract – Corporate Social Responsibility (CSR) reports constitute a relatively new form of corporate disclosure used by companies to present their values and philosophy with respect to socially relevant themes on which they may have an impact, mainly the environment, the community and employees. Companies thus publish CSR reports to communicate with a variety of stakeholders and provide information about their sustainability initiatives, with the ultimate aim of building, reinforcing, and promoting their corporate image. Personalisation plays an important role in the discursive construction of identity and in the definition of relationships between social actors. The personification of the company – obtained through 1st person plural deixis within corporate reports – is a very powerful rhetorical tool to convey a collective subject which takes responsibility for the actions and results it is giving account of, indicating and enacting a specific relationship with the reader. As a sociopragmatic item, however, it is largely language/culture-dependent, and thus represents an interesting locus to observe the impact of translation strategies on the meaning conveyed to the target audience. This paper sets out to analyse how CSR reports translated into English from Italian compare – as regards personalisation – with reports originally produced in English, in order to detect differences in the way corporate identity is construed and conveyed. The study is based on a bilingual corpus which includes translated English reports and their Italian source texts, as well as comparable originals in English and Italian. Corroborating previous research conducted on similar genres, the study shows that (im)personalisation patterns are considerably different in original and translated English CSR reports, largely due to a tendency for the latter to reproduce Italian conventions in this form of specialised discourse.

Keywords: corporate discourse; sociopragmatics; (im)personalisation; translation; corpus linguistics.

1. Introduction to Corporate Social Responsibility

The notion of Corporate Social Responsibility (CSR) refers to action and policies adopted by companies to take responsibility for their business' impact on society and the environment. In the words of the European Commission (2017), companies can become socially responsible by "integrating social, environmental, ethical, consumer, and human rights concerns into their business strategy and operations" and in their interaction with stakeholders.[1] This commitment should not only fulfil legal obligations and regulatory requirements: the crucial and innovative aspect of CSR is that companies are expected to implement actions and policies in the area of corporate sustainability on a voluntary basis (COM(2001) 366 final),[2] in order to show how their activity reflects ethical governance and how it tangibly contributes to the common good.

To inform stakeholders of their 'sustainable' activities, companies issue social responsibility reports. CSR reports are key documents where data on corporate sustainability is disclosed and described in detail: they provide "information relating to a corporation's activities, aspirations and public image with regard to environmental, community, employee and consumer issues […] such as energy usage, equal opportunities, fair trade, corporate governance and the like" (Gray *et al*. 2001, p. 329). CSR reports have become standard practice for corporate communication on sustainability issues, especially among multinationals, and nowadays they are almost entirely published online, on the company's internet website (Yu, Bondi 2017). The recipients of this information range from potential investors to shareholders, from the general public to selected activist groups operating in the relevant community and to anyone else who may be interested "in the social performance of the company as a predictor of its financial performance" (Fasan 2013, p. 45). Fasan's remarks suggest that investors take social and environmental issues into consideration when making investment decisions, thus highlighting the financial potential of CSR. Over the years, the business payoffs of companies' sustainability actions have been highlighted, revealing the financial advantages of 'going sustainable'. The underlying rationale is that companies that do some good to society and to the environment become justified and legitimated to operate due to their social utility (Fuoli 2012, p. 4; see also Fuoli 2017; Bondi 2016b).

[1] http://ec.europa.eu/growth/industry/corporate-social-responsibility_it

[2] Green Paper of the Commission of the European Communities of 18 July 2001, "Promoting a European framework for Corporate Social Responsibility" (COM(2001) 366 final). See also the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions of 25 October 2011, "A renewed EU strategy 2011-14 for Corporate Social Responsibility" (COM(2011) 681 final).

Corporate commitment towards indisputable ideological and moral principles – such as environmental protection, equal employment, fair trade, decent work conditions, corporate ethics etc. – provides a company with appealing and convincing reasons that justify its activity and fortify its legitimacy framework (Cap 2005, p. 13). In particular, a twofold legitimation process emerges: 'legitimation through idea', calling on the undebatable good values and principles evoked in CSR reports, and 'legitimation through procedure', whereby the concrete beneficial actions described in CSR reports sanction de facto the company's social utility (Wodak, Weiss 2005, p. 131). Against this background, legitimation of corporate operations is achieved simultaneously by 'moral evaluation', as the company's action is based on shared and indisputable values in line with society's (assumed) wants and needs, and by 'rationalisation', as the company's action is justified by its social utility and by its beneficial effects towards the common good (Van Leeuwen 2007, p. 92; see also Van Leeuwen 1996; Cap 2005). Hence, corporate operations appear not only morally legitimate but also justified on a utility principle. In other words, the general public does not only perceive the company as reliable and worthy of being endorsed but it also becomes persuaded that that specific company needs to exist – over other less virtuous companies – in order to promote a healthy change in corporate culture and to contribute to a shift towards sustainable production (Higgins, Walker 2012). This, in turn, translates into a definite financial advantage: the public chooses to buy from or invest in that specific company not just to fulfil consumer needs, but also to ultimately help 'a good cause'.

2. Study background and purpose

Effectively communicating the company's responsible approach to its stakeholders is crucial for the company to gain a financial advantage. A vast body of literature analysing general and specific rhetorical and discourse features of non-translated CSR reports is available (Catenaccio 2011a, 2011b, 2012a, 2012b, 2013, 2014, 2017; Catenaccio, Degano 2011; Bhatia 2012; Aiezza 2015; Bondi 2016b; Fuoli 2012, 2017; Higgins, Walker 2012; Skulstad 2008; Malavasi 2011, 2012, 2017; Ocler 2009), or adopting a multimodal approach to investigate their semiotic dimension (Höllerer *et al*. 2013; Maier 2014; Catellani 2015). Research has also been carried out in a contrastive perspective, comparing discourse features of CSR reports across languages (e.g. Russian/English, Italian/English, Italian/Chinese/English etc.) (Bondi, Yu 2017; Yu, Bondi 2017; Bondi 2016a; Bashtovaya 2014; Bhatia 2012). Specifically on the language combination English/Italian, contrastive research has mostly focused on English and Italian corporate texts not (necessarily) related to CSR issues – namely annual reports, CEO letters and business correspondence (Garzone 2004; Giannoni 2001, 2014; Vergaro 2002, 2004, 2005).

The purpose of this study is, therefore, to contribute to filling an important gap in the body of academic research on CSR discourse, that is the analysis of translated CSR texts. A large number of translated reports can be found online and, when full reports are not available, multilingual translated summaries are usually available. Hence, a widespread translation trend can be observed in CSR online communication – possibly due to the social relevance and business impact of the sustainability issues mentioned earlier – a trend showing how corporations aim at reaching the widest possible audience worldwide in their communication on CSR issues.

In terms of the specific focus of our analysis, we have decided to concentrate on strategies of personalisation, more specifically on 1st person plural deixis. As previously pointed out by Fuoli (2012, p. 2), most discourse investigations of CSR reports have primarily focused on their ideational function, while their interpersonal function – that is how reporting companies construct their identity and relationship with their interlocutors – has not been the object of systematic investigation. Personalisation, in particular, has been found to play an important role in the discursive construction of identity (Fairclough 2001) and in the definition of relationships between social actors (Van Leeuwen 1996; see also Wodak *et al*. 2009; Wodak 2012). Van Leeuwen (1996, p. 59) describes personalisation as a crucial element in providing social actors with the ‘human’ feature: it is ranked among the most prominent – and effective – communicative strategies used to endow an actor with a comforting and appealing ‘human side’ (Caliendo, Magistro 2009).

A few studies of similar corporate texts – namely annual reports – have indeed looked at the representation of corporate identity conveyed through personalisation. Garzone’s (2004) study on the use of personal pronouns within *Letters to shareholders* – arguably a well-defined sub-genre within annual reports – is a case in point. Garzone found that corporate statements written in English by Italian companies were characterised by a lower frequency of the pronoun *we* and a higher occurrence of impersonal and passive forms than texts produced by multinationals located in other countries, arguably due to interference from specialised discourse conventions in the Italian language.

Among the different strategies to obtain a ‘personalised’ effect, the use of *we*-references is indeed crucial in conveying the image and role of the first turn taker in the dialogue that the company establishes with the stakeholder/investor/client or the wider public in general. Fairclough (1995, p. 145) notes how *we*-references give an impression of cooperative collegiality and confer a dialogic tone, presupposing a *you*-listener (see also Bargiela-Chiappini, Harris 1997). Commenting on corporate annual reports, Garzone adds that 1st person (plural) deixis is used to convey a collective subject, confronting addressees with a corporate 'persona':

> a textual self which takes upon itself responsibility for the actions and results for which s(he) is giving account and for the decisions and strategies illustrated in the text. In other words, recourse to first person pronouns as subjects reflects a form of ethical commitment for the writer; it is in the service of ethos. (Garzone 2004, p. 236)

The 'personification' of the company is therefore a very powerful rhetorical tool that not only conveys closeness with the reader, but de facto enacts it. Regardless of whether this closeness is wanted by the reader, by discursively establishing such closeness the company already makes it a fact. The reader can only acknowledge this closeness and embrace it in order to participate in the dialogue with the company.

2.1. Theoretical framework

The theoretical assumptions underlying our study lie in discourse analysis and translation studies.

We first assume that pragmalinguistic elements carry culturally-loaded information: they convey the speaker's perception of appropriate language and acceptable/desirable communication in a given context (Leech 1983). Bondi (2016a, p. 168) highlights a very interesting aspect on the notion of culture by noting that cross-cultural discourse perspectives on CSR are mostly interested in corporate culture itself over 'national culture'. She adds that "because of the global dimension of both sustainability issues and business in general, disclosure documents and CSR reports in particular are […] clearly related to the global and international contexts, with a clear dominance of international guidelines". In other words, CSR reports have become a globally established and standardised genre with remarkable similarities in different languages, "suggesting that the communicative purposes of CSR reports are recognised by different cultures" (Yu, Bondi 2017). While acknowledging the global dimension of corporate 'culture' – featuring a worldwide discourse community of corporate actors – in this paper we refer to 'culture' as the communicative habits of a speech community embracing specific behavioural conventions, discursive expectations and norms of social appropriateness (Kramsch 2003, p. 6), thus not limited to corporate or business settings. More specifically, we will focus on texts produced by a community of Italian-language users and on their respective translations into English.

We then adopt a critical approach, as we intend to discuss the impact of translators' choices on the sociopragmatic perceptions of the speech community receiving these translations. In particular, we assume that discourse is socially performative as it plays an important role in shaping social identities and reality configurations (Fairclough 1993, p. 139). Based on the postulations of Critical Discourse Analysis, discourse is a form of social practice that can produce actual change in society by affecting (the construction of) social relationships and identities, as well as frameworks of knowledge (Fairclough, Wodak 1997, p. 258). Language can be used to affect society by portraying a desirable scenario governed by desirable relations; this scenario does not necessarily represent the actual status quo, rather the scenario and relations that the company intends to establish – allegedly – to the benefit of the readers and the broader world community. The translator, therefore, becomes a powerful agent for cultural change: a translator's decision on how to render pragmalinguistic elements of the source text can have an impact on the perceptions and communicative habits of the receivers, thus on the target culture(s) (Venuti 1995).

What is 'faithfully rendered' from a pragmalinguistic perspective, for example, may be perceived as 'not appropriate' by – or frustrate the expectations of – the target audience from a sociopragmatic or cultural perspective. In other words, using in the target language the resources provided by the source language to convey a particular illocution may not result in an illocution which is consistent with the initial one (Leech 1983, p. 11). Hence, transferring the same pragmalinguistic features may prove to be sociopragmatically and culturally unsuitable for – or simply 'unfamiliar' to – the target audience, as it may generate pragmatic effects that deviate from the communicative habits and expectations of the target speech community (Magistro 2013).

Quoting Toury's law of interference (1995, p. 275), when "phenomena pertaining to the make-up of the source text tend to be transferred to the target text" – determining an over- or under-representation of features that do exist in the TL but with different frequencies – the result may be target language discourse that is grammatically correct but unnatural. As pointed out by Mauranen (2004), frequency deviations from native TL usage may virtually go unnoticed by ordinary readers and be detected only by means of large-scale frequency comparisons. In this regard, the application of corpus linguistics methods to translation studies – and most notably the extensive body of research based on monolingual comparable corpora conducted over the last twenty years – have proved fundamental for pinpointing differences between native and translated language at several linguistic levels, from syntax to lexical density and variety.[3]

[3] Monolingual comparable corpora are corpora of native and translated texts in the same language. The monolingual comparable corpus approach became mainstream especially after Baker (1993) suggested that the comparison of translated and non-translated texts could unveil "universal features of translation", i.e. features typically occurring in translated texts rather than in original

3. The study: research questions, corpus, method and results

The study focuses on the use of 1st person plural deixis to build corporate identity in English and Italian original CSR reports and on how personalisation – as a discursive strategy – is dealt with in Italian-to-English translations. More particularly, by examining how companies are portrayed and which type of writer-reader relationship (i.e. personal/impersonal) is established in the two languages, the study sets out to investigate whether personalisation patterns observed in translated English CSR reports are different from those found in original English CSR reports, and to what extent they are affected by Italian socio-pragmatic habits. The ultimate purpose is to discuss the consequences translation choices might have on the perception of corporate image conveyed through translated reports.

3.1. Corpus description

In order to answer the above questions, it was necessary to assemble an *ad-hoc* bilingual corpus which included original Italian and English reports as well as translated English reports. Such a composite resource can be queried according to different perspectives, most notably a contrastive perspective (i.e. original English reports vs. original Italian reports) and a monolingual comparable perspective (i.e. original English reports vs. translated English reports). The Italian source texts corresponding to the translated English reports were also included, even though the specific aims of this study do not require a proper parallel component. In fact, the study does not aim to extract and analyse specific translation examples, rather to observe the possible overall impact of the Italian language on the translated subcorpus.

The data for this study come from a corpus of 251 CSR reports published between 2012 and 2017 by companies operating in a variety of sectors, mainly energy, financial services, food and beverage, and construction. In order to enhance their comparability, the texts were retrieved through an existing online repository – the *Sustainability Disclosure Database*[4] – which at the time of writing contains over 48,000 CSR reports produced by more than 12,000 organisations worldwide. The repository is maintained by the Global Reporting Initiative, a non-governmental non-profit international organisation which

utterances as a result of the process of translation. “Universal differences between translations and comparable non-translated texts” would later be defined as *T-universals* by Chesterman, in opposition to *S-universals*, that is “universal differences between translations and their source texts” (Chesterman 2004). See e.g. Laviosa (2002), Zanettin (2012) for interesting reviews of studies based on monolingual comparable corpora.

[4] http://database.globalreporting.org/

provides internationally recognised standards and guidelines for sustainability reporting.[5] Most CSR reports included in the database are thus GRI-compliant, although the repository also provides access to reports that are not GRI-based. Overall the corpus totals 8.5 million tokens, as shown in Table 1 (the -EN and -IT affixes indicate the language of the texts – English and Italian respectively – and the -ST, -TT, -REF suffixes indicate source, target and reference texts).

Sub-corpus	**Language**	**Description**	**Texts**	**Tokens**
CSR-EN-REF	English	Original English reports	82	2,308,955
CSR-EN-TT	English	Translated English reports, corresponding to CSR-IT-ST	59	2,196,175
CSR-IT-ST	Italian	Original Italian reports, source texts for CSR-EN-TT	59	2,234,907
CSR-IT-REF	Italian	Original Italian reports	51	1,739,066

Table 1
Corpus details.

Relevant parallel CSR reports were retrieved by applying the country filter available in the repository search interface, starting from the assumption that business organisations recorded as having their headquarters in Italy are likely to have their CSR reports originally written in Italian and subsequently translated into English. As this was found to be overtly stated only in a few reports (8/59), we contacted all the selected companies to ascertain that the English version was truly a translation of the Italian version. About half of the companies (27/59) replied to our query and confirmed that their CSR report was originally written in Italian. In addition, in over two thirds of the translated English reports it is explicitly stated that the attached independent auditor's report is a translation of the original Italian version, which would prevail in case of dispute.[6] This statement, per se, indirectly confirms that the audited texts (i.e. our sustainability reports) were originally written in Italian.[7] Only for

[5] http://www.globalreporting.org. GRI standards are not legally binding, still they are widely adopted on a voluntary basis across the corporate community worldwide, including 74% of the world's largest 250 companies (source: https://www.globalreporting.org/information/sustainability-reporting/Pages/gri-standards.aspx). These guidelines provide indications of *what* should be included in the reports, but do not say much about *how* the information should be articulated (Catenaccio 2011a, p. 72).

[6] For the sake of example, the English versions of the independent auditors' reports include statements like "Translation from the Italian original which remains the definitive version", "This report has been translated from the original, which was issued in Italian, solely for the convenience of international readers" and "Signed on the original" (showing no signature on the English version).

[7] As a matter of fact, the independent auditor officially certifies that the relevant sustainability report is prepared in compliance with the latest Sustainability Reporting Guidelines defined by the Global Reporting Initiative (GRI), with all due legal implications of an official recognition. Such a certification can only be granted if the certifier understands the content (hence the language) of

one fifth of the companies (12/59) is no specification provided on the language of the source text. In these cases we embraced the assumptions that a company a) whose headquarters and registered offices are based in Italy, and whose management and founders are Italian, b) whose website is available mostly or entirely in Italian, and c) whose sustainability report largely refers to compliance with Italian legislation and relations with Italian public institutions, local associations (e.g. Cral) and contests (e.g. Premio Compasso d'Oro) is a company that is very likely to have strong ties with the Italian community and, therefore, is very likely to circulate key reports impacting the local context in the community's language – Italian, that is.

The same assumptions hold true for the selection of original texts. The English comparable component includes CSR reports published by companies with headquarters in several countries where English is an L1 – mainly UK and US, but also Ireland, Australia, New Zealand and South Africa – while the comparable corpus of Italian contains CSR reports produced by Italian companies (see the above criteria) for which no translation was available. The latter subcorpus was used to obtain information about personalisation patterns commonly found in Italian CSR reports, against which to analyse patterns featured in Italian STs.

The reports were downloaded in PDF format from either the repository or the official websites of the companies involved, then converted into raw text format using unsupervised (fully automated) file conversion software.

3.2. Method

The research procedure primarily consisted of three steps. The first step involved singling out all possible *we*-references in English and Italian, i.e. all linguistic items instantiating the 1st person plural which contribute to the personalisation of companies as collective subjects (*we*) and can be exploited for the rhetorical purpose of modulating the company's involvement (compared to the use of passive and interpersonal forms, which convey a lower degree of involvement). The two languages involved are characterised by morphosyntactic differences: while considering personal pronouns (*we*, *us*, *ours* and *ourselves*) and the possessive adjective *our* was sufficient for English, for Italian – which is an inflected pro-drop language – it was necessary to also include inflected forms for verbs and possessive determiners. As this step was crucial for the study, and the quality of the available Italian POS-tagging was not deemed satisfactory, a list of suffixes which are indicative of 1st person

the document to be certified, otherwise a necessary precondition for the certification would not be met. In our case, if Italian is the language of the original certification (i.e. of the original auditor's report), it can be assumed that the audited document was also drafted in Italian.

plural forms across all verb tenses was compiled and used in wildcard searches (see Table 2).

Language	Search words
English	we | us | our | ours | ourselves
Italian	noi | nostr* | ci *iamo | *remo | *vamo | *mmo | *ssimo | *rci

Table 2
Search words.

As a second step, *we*-references were extracted from the different subcorpora – and their frequency calculated – by retrieving concordances for the search words listed in Table 2, using the *AntConc* software (Anthony 2014). The concordance function was chosen (over simple word counts) because it allowed for the subsequent manual pruning of irrelevant hits. These mainly included words ending in one of the specified suffixes but not relevant for the analysis, or occurrences of words from the search list which were retrieved from parts of the reports other than the actual body (e.g. boilerplate text which had not been properly removed by the file conversion tool).

The third stage consisted in comparing the normalised frequencies of *we*-references across subcorpora. As a first step, normalised frequencies in the two English subcorpora were computed in order to detect quantitative differences between non-translated and translated English reports. The statistical significance of the observed frequency differences was calculated using the log-likelihood test (as proposed, among others, by Rayson, Garside 2000);[8] the %DIFF effect-size metric (Gabrielatos, Marchi 2011) was also used to measure the size (as a proportion) of the difference between the normalised frequencies in the two corpora. The English data were then compared to data extracted from the Italian subcorpora to assess whether the frequencies observed in translated reports could somehow be linked to Italian personalisation patterns and, ultimately, the translation process.

[8] Log-likelihood values were calculated using the online tool developed and made available by Lancaster University at http://ucrel.lancs.ac.uk/llwizard.html. The largest LL values indicate the words having the most significant relative frequency difference between the two corpora. The chi-square test applied to the same frequency data further cross-checked the validity of the results.

3.3. Results

For the monolingual comparable comparison of *we*-references across original and translated reports, frequencies of the English items in Table 2 were calculated in the two English subcorpora (see Table 3).

	CSR-EN-TT		**CSR-EN-REF**		**Log-Likelihood**	**%DIFF**
	f - raw	**f - ptw**	**f - raw**	**f - ptw**		
we	3,037	1.38	22,652	9.81	15977.27	-85.90
our	3,757	1.71	36,108	15.64	28774.15	-89.06
us	457	0.21	2,225	0.96	1182.45	-78.41
ours	12	0.01	21	0.01	2.06	-39.92
ourselves	37	0.02	129	0.06	49.48	-69.84
Total	7,300	3.32	61,135	26.48		

Table 3
We-references in the two English subcorpora.

It can be observed that virtually all 1st person plural references are more frequent in original English reports than in translated reports. The difference is particularly striking for the subject pronoun *we* (9.81 ptw in CSR-EN-REF vs. 1.38 ptw in CSR-EN-TT, LL=15977.27, %DIFF=-85.90)[9] and the possessive adjective *our* (15.64 vs. 1.71, LL=28774.15, %DIFF=-89.06), which incidentally represent the two most frequent *we*-references in both subcorpora, with a large impact over overall frequencies (26.48 vs. 3.32). All frequency differences are statistically significant ($p<0.0001$) with the exception of the pronoun *ours*. Translated CSR reports thus appear to be characterised by a much lower degree of personalisation than original English reports, in line with previous findings by Garzone (2004) on similar corporate texts (see Section 2).

In order to determine whether this linguistic feature can be recognised as a consequence of the texts being translated from Italian, the frequency of 1st person plural references in the corresponding Italian source texts was then taken into account. Relevant figures are reproduced in Table 4, together with data from the Italian reference subcorpus. The total normalised frequency of 1st person plural references in the Italian source texts is 2.49 per thousand words, only slightly higher than the normalised frequency observed in the Italian reference corpus (2.09). This means that the degree of personalisation of Italian STs is in line with what is commonly found in the CSR report genre

[9] Negative %DIFF values indicate that the relevant word has a higher normalised frequency in the reference corpus.

for Italian, namely a much less extensive use of 1st person plural references than in original English CSR reports.

	CSR-IT-ST		**CSR-IT-REF**	
	f - raw	**f - ptw**	**f - raw**	**f - ptw**
noi	174	0.08	150	0.09
nostr*	2,602	1.16	1,549	0.89
VERB	2,540	1.14	1,756	1.01
ci	247	0.11	176	0.10
Total	**5,563**	**2.49**	**3,631**	**2.09**

Table 4
Raw and normalised frequencies of *we*-references in the two Italian subcorpora.

Joining data from Tables 3 and 4, it becomes apparent that the normalised frequency of *we*-references in English translated CSR reports (3.32 words per thousand words) is much closer to their frequency in the corresponding Italian source texts (2.49) than in comparable non-translated English reports (26.48). It can thus be safely argued – even without extracting actual parallel concordances – that the personalisation patterns of the Italian source texts tend to be largely reproduced in translated texts.

Further investigations on *we* verb collocates (Castagnoli, Magistro 2018) and the manual analysis of concordances showed that the incidence of *we*-inclusive forms – as exemplified by (1) below, where the underlined *we* arguably refers to society as a whole – is virtually irrelevant across all the different corpus components. In both English and Italian reports, *we*-references are mostly exclusive, that is they are used only to refer to the companies authoring the reports, without involving the readers in the creation of a united entity sharing the same perspective (see examples 2-4).

(1) Your contribution will help us continue to improve outcomes for our industry and our communities so we can address the challenges we face together. (CSR-EN-REF)
(2) We actively look for opportunities to reduce waste. (CSR-EN-REF)
(3) We are also working towards involving our employees in our new sustainability agenda, and we aim to improve transparency across the organisation and embed sustainability into our daily routines. (CSR-EN-TT)
(4) Promuovere costantemente questi temi è per noi una priorità. (CSR-IT-ST)

3.4. Discussion

The data presented in Section 3.3 show that personalisation in Italian CSR reports, as given by *we*-references, is much lower than in original English CSR reports. This is possibly an indication that Italian textual conventions in specialised discourse – including corporate discourse about CSR – still require a certain degree of formality, which is assumed to convey an impression of professional competence and objectivity (Garzone 2004). The low occurrence of 1st person deixis in Italian texts might indicate that, as further observed by Garzone (2004, p. 337), Italian companies may still not have elaborated a more empathic communicative style due to the fact that Italy has only recently recorded a real peak of interest in corporate investment on the part of ordinary people. Although the CSR report can now be considered as a highly standardised genre in terms of disclosure contents, textual structure and high-level rhetorical moves (see Section 2.1), culture-bound preferences as regards the use of specific pragmalinguistic features – such as personalisation – still determine divergencies in this form of specialised discourse across languages.

These differences are crucial in the translation process. Target-oriented approaches to translation – which prioritise target text *acceptability* (Toury 1995), i.e. adherence to target language norms and ultimately target reader's expectations – are indisputably the norm in current specialised translation. If the more extensive usage of *we*-forms in original English reports reflects conventions for the CSR genre in English-speaking countries – and translated reports in the corpus turn out to be more similar to their Italian source texts in terms of the incidence of 1st person deixis – one is left to wonder to what extent CSR reports translated from Italian can be considered pragmatically 'acceptable'. By mimicking the pragmalinguistic elements featured in the Italian source text, without performing any major domesticating intervention with respect to the use of 1st person plural deixis (as suggested by the comparison with the English reference corpus), translators produce translated reports that are characterised by a more prominent distancing component compared to English reference texts. This might be perceived as sociopragmatically unsuitable by the target audience.

However, defining a target audience in this case is not a trivial task. When translating for the international corporate discourse community, which uses English as a lingua franca, it may not be possible for translators to determine what it means to "achieve the greatest possible correspondence, referentially and *pragmatically*" (Newmark 1998, p. 23, emphasis added). Stakeholders worldwide are inevitably imbued with conventions and expectations of their own cultures/speech communities (see Section 2), and what may be acceptable for target readers of a given speech community might at the same time frustrate other stakeholders' discursive expectations. Research on audience reception would be needed in order to ascertain whether the production of an unusual or 'marked' sociopragmatic scenario in translated texts can have repercussions on corporate image.

As for all product-oriented research, it is not possible to determine whether the features observed in translated texts are due to deliberate stylistic choices or rather the result of interference from the sociopragmatics of source texts. Translators might be willing to preserve at least 'some' degree of impersonalisation, either to reproduce more 'faithfully' the corporate image conveyed by Italian companies in Italian originals (in a more source-oriented approach), or to comply with an alleged reluctance of Italian companies to adopt a more personalised communicative style.

To try and overcome this limitation, we attempted to gather some information on the translation process from the companies involved in our research (see also section 3.1). We asked them who carried out the translation (i.e. internal staff, external translators, translation agencies) and whether, to their knowledge, the translator was a native-speaker of English. We also asked if a translation brief with the company's stylistic preferences was provided. About half the respondents confirmed that stylistic or terminological directions were provided or that the translator had already collaborated with the company and, therefore, was familiar with the company's preferences. As it is normally the case in the translation industry, we were not given the names of the translators/translation agencies involved in the translation of the reports – mostly for privacy reasons. It was therefore impossible to contact them and gather information on the rationales for their choices. However, even if we had been given the names of individual translators and had had the opportunity to consult them for a retrospective analysis of their choices, a major disadvantage comes to mind regarding the retrospective practice – a downside that could even distort any results. Research has shown that interviewing a translator can provide valuable insights on the process and cognitive mechanisms of translation when the interview ideally takes place immediately after the translation task, so as to minimise memory issues and cued recalls:

> A disadvantage of retrospection is that it generally does not allow complete recall of the information, especially in longer tasks, for instance, translation. […] The data is based on the participants' recall, and reports may be distorted for many different reasons. (Englund Dimitrova, Tiselius 2014, p. 177)

Our reports and their respective translations were published between 2012 and 2017, that is at least a year – and often several years – before the time this paper was written, thereby posing a major problem in terms of the immediacy of the translator's consultation.[10] Moreover, this discussion refers to the unusual scenario of a single translator performing the task, whereas it is widely acknowledged in translation research and practice that translating is often a collaborative task involving different subjects. For example, more than one translator may have worked on the text, a reviser and/or a proofreader may have made changes to the text at a later stage, or the company's internal staff may have intervened to amend the translation (this case was actually mentioned by many of the companies that we have contacted).

Although process-oriented research is interesting and challenging for the wider field of translation studies, it is not the focus of this corpus-based study. No doubt the triangulation of corpus data with some ethnographic data (from basic information about translators to structured interviews about possible motivations for translation strategies) would represent an advance in research circumstances that allow for more immediate consultation of the subjects involved in the translation task (e.g. translation students, in-field ethnographic projects, interpreting tasks etc.). Unfortunately this type of investigation cannot be realistically conducted in the context being considered: collecting interviews or other ethnographic data so long after the actual translation took place would have exposed the study to the material risk of distorting or even cueing results.

4. Final remarks

The pilot study described in this paper has provided some evidence that the corporate image conveyed in English CSR reports translated from Italian is substantially different from the one conveyed through comparable originals, in that the underuse of *we*-references in translated reports – probably mirroring source-language sociopragmatic conventions – results in less personal, less proximising and possibly more formal corporate disclosures. In order to obtain a more comprehensive picture of personalisation patterns, however, other forms need to be taken into consideration – such as *I*- and especially *you*-references, possibly addressing readers directly and enhancing their involvement (in lieu of *we*-inclusive forms, as mentioned in Section 3.3). These need to be coupled with an analysis of strategies which avoid personalisation

[10] When a translation task occurs in a time that is so distant in the past, obviously the (high) risk is that facts may be remarkably blurred in the translator's mind, thus forcing them to provide an answer to the researcher by 'fabricating' memories or by simply making logical assumptions on the reasons for their translation choices that may not reflect their line of reasoning when the task actually took place. In other words, whereas the locution – the actual translation – has remained unaltered since the 2012-2017 period, the illocution – the translator's intentions – may be perceived differently now due to a number of reasons (e.g. the translator is now more experienced, the translator has gained access to new information, new technological tools have become available, their work has been subject to revision etc.).

in corporate reports, such as using 3rd person references, agentless passives and impersonal forms. Further work will include observing how lexical items like *group*, *company* or *business* are exploited to this end. Collocational profiles for the subject pronoun *we* and the possessive adjective *our* will also be examined (Castagnoli, Magistro 2018) in order to derive further information about the differences in the way corporate image is construed in original and translated reports.

Bionotes: Sara Castagnoli is Assistant Professor of English Language and Translation at the University of Macerata. She holds a Ph.D. in Linguistics from the University of Pisa and her research focuses on translation studies, learner translation, LSP, terminology and lexicography, especially from a corpus-based perspective. Her recent publications include *Investigating Trainee Translators' Contrastive Pragmalinguistic Competence: A Corpus-based Analysis of Interclausal Linkage in Learner Translations* (2016), *Sfide cross-culturali nella traduzione specializzata* (2016) and *Corpora worth creating: a pilot study on telephone interpreting* (2017).
Elena Magistro is Adjunct Professor of Specialized Translation at the University of Bologna-Forlì. She holds a Ph.D. from the University of London and her research interests lie in translation, discourse analysis, pragmatics, linguistic politeness and European identity. She also collaborates as a translator and language consultant with European Union institutions and agencies. Her publications include: *The discursive technology of Europeans' involvement: EU* culture *and community of practice* (2014), *The challenges of 'translating' polite discourse for the EU multilingual community* (2013) and *National face and national-face threatening acts* (2011).

Authors' addresses: sara.castagnoli@unimc.it, elena.magistro2@unibo.it.

References

Aiezza M.C. 2015, *"We may face the risks" … "risks that could adversely affect our face." A corpus-assisted discourse analysis of modality markers in CSR reports*, in "Studies in Communication Sciences" 15 [1], pp. 68-76.

Anthony L. 2014, *AntConc (3.2.4)* [Computer Software], Waseda University, Tokyo. http://www.laurenceanthony.net/software (30.4.2018).

Baker M. 1993, *Corpus Linguistics and Translation Studies – Implications and Applications*, in Baker M., Francis G. and Tognini-Bonelli E. (eds.), *Text and Technology – In Honour of John Sinclair*, John Benjamins, Amsterdam/Philadelphia, pp. 233-250.

Bargiela-Chiappini F. and Harris S. 1997, *Managing Language: The Discourse of Corporate Meetings*, John Benjamins, Amsterdam/Philadelphia.

Bashtovaya V. 2014, *CSR reporting in the United States and Russia*, in "Social Responsibility Journal" 10 [1], pp. 68-84.

Bhatia A. 2012, *The corporate social responsibility report: The hybridization of a 'confused' genre (2007–2011)*, in "IEEE Transactions on Professional Communication" 55 [3], pp. 221-238.

Bondi M. and Yu D. 2017, *"I highly commend its efforts to ensure power supply": Exploring the pragmatics of textual voices in Chinese and English CSR reports*, in Janebová M., Lapshinova-Koltunski E. and Martinková M. (eds.), *Contrasting English and Other Languages through Corpora*, Cambridge Scholars Publishing, Newcastle, pp. 218-247.

Bondi M. 2016a, *CSR reports in English and Italian: Focus on generic structure and importance markers*, in Garzone G., Heaney D. and Riboni G. (eds.), *Language for Specific Purposes. Research and Translation across Cultures and Media*, Cambridge Scholars Publishing, Newcastle, pp. 168-199.

Bondi M. 2016b, *The future in reports: Prediction commitment and legitimization in corporate social responsibility (CSR)*, in "Pragmatics and Society" 7 [1], pp. 57-81.

Caliendo G. and Magistro E. 2009, *The human face of the European Union: A critical study*, in "CADAAD Journal" 3 [2], pp. 176-202.

Cap P. 2005, *Language and legitimization: Developments in the proximization model of political discourse analysis*, in "Lodz Papers in Pragmatics" 1, pp. 7-36.

Castagnoli S. and Magistro E. 2018, *Reshaping corporate image through translation. A corpus-assisted study*, paper presented at the Corpora and Discourse (CAD 2018) international conference held at Lancaster University, Lancaster (UK), 22-24 June 2018.

Catellani A. 2015, *Visual aspects of CSR reports: A semiotic and chronological case analysis*, in De Melo A., Somerville I. and Gonçalves G. (eds.), *Organisational & Strategic Communication Research: European Perspectives II*. CECS – Centro de Estudos de Comunicação e Sociedade, Universidade do Minho, Braga, pp. 129-149.

Catenaccio P. 2011a, *Social and environmental reports. A diachronic perspective on an emerging genre*, in Garzone G. and Gotti M. (eds.), *Discourse, Communication and the Enterprise. Genres and Trends*, Peter Lang, Bern, pp. 169-192.

Catenaccio P. 2011b, *Green advertising and corporate CSR communication: Hybrid discourses, mutant genres*, in Sarangi S., Polese V. and Caliendo G. (eds.), *Genre(s) on the Move. Hybridization and Discourse Change in Specialized Communication*, ESI, Napoli, pp. 353-372.

Catenaccio P. 2012a, *Understanding CSR Discourse. Insights from Linguistics and Discourse Analysis*, Arcipelago, Milano.
Catenaccio P. 2012b, *The 'value-orientation' of business discourse: The role of corporate social responsibility in business communication. A case study*, in Heynderickx P., Dieltjens S., Jacobs G., Gillaerts P. and de Groot E. (eds.), *The Language Factor in International Business: New Perspectives on Research, Teaching and Practice*, Peter Lang, Bern, pp. 247-265.
Catenaccio P. 2013, *The discursive encoding of changing business values in CSR reports: A corpus-based investigation*, in Poppi F. and Cheng W. (eds.), *The Three Waves of Globalization*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 56-76.
Catenaccio P. 2014, *The evolution of business discourse and the emergence of the corporate social responsibility paradigm: An investigation of CSR reports*, in Evangelisti Allori P., Bateman J.A. and Bhatia V.K. (eds.), *Evolution in Genre. Emergence, Variation, Multimodality*, Peter Lang, Bern, pp. 19-44.
Catenaccio P. 2017, *Markers of trust: Epistemic adverbs of certainty and restrictive adverbs in CSR reports*, in Salvi R. and Turnbull J. (eds.), *The Discursive Construal of Trust in the Dynamics of Knowledge Diffusion*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 88-107.
Catenaccio P. and Degano C. 2011, *Corporate social responsibility as a key to the representation of corporate identity: The case of Novartis*, in Evangelisti Allori P., Bateman J.A. and Bhatia V.K. (eds.), *Evolution in Genre. Emergence, Variation, Multimodality*, Peter Lang, Bern, pp. 79-102.
Chesterman A. 2004, *Beyond the particular*, in Mauranen A. and Kujamäki P. (eds.), *Translation Universals. Do they exist?*, John Benjamins, Amsterdam/Philadelphia, pp. 33-49.
Englund Dimitrova B. and Tiselius E. 2014, *Retrospection in interpreting and translation: explaining the process?*, in "MonTI" Special Issue – Minding Translation, pp. 177-200.
Fairclough N. 1993, *Critical discourse analysis and the marketization of public discourse: The universities*, in "Discourse and Society" 4 [2], pp. 133-168.
Fairclough N. 1995, *Critical Discourse Analysis: The Critical Study of Language*, Longman, London.
Fairclough N. 2001, *Language and power* (2nd edition), Longman, London.
Fairclough N. and Wodak R. 1997, *Critical discourse analysis*, in Van Dijk T.A. (ed.), *Discourse as Social Interaction. Discourse Studies: A Multidisciplinary Introduction* (Volume 2), Sage Publications, London, pp. 258-284.
Fasan M. 2013, *Annual reports, sustainability reports and integrated reports: Trends in corporate disclosure*, in Busco C., Frigo M.L, Riccaboni A. and Quattrone P. (eds.), *Integrated Reporting: Concepts and Cases that Redefine Corporate Accountability*, Springer, Berlin, pp. 41-58.
Fuoli M. 2012, *Assessing social responsibility: A quantitative analysis of appraisal in BP's and IKEA's social reports*, in "Discourse and Communication" 6 [1], pp. 55-81.
Fuoli M. 2017, *Building a trustworthy corporate identity: A corpus-based analysis of stance in annual and corporate social responsibility reports*, in "Applied Linguistics", pp. 1-41.
Gabrielatos C. and Marchi A. 2011, *Keyness: Matching metrics to definitions*, invited presentation at *Corpus Linguistics in the South 1*, University of Portsmouth, 5 November 2011.

Garzone G. 2004, *Annual company reports and CEO's letters: Discoursal features and cultural markedness*, in Candlin C.N. and Gotti M. (eds.), *Intercultural Aspects of Specialized Communication*, Peter Lang, Bern, pp. 311-342.

Giannoni D.S. 2001, *Politeness in English and Italian business discourse*, in Bettoni C. Zampolli A. and Zorzi D. (eds.), *Atti del 2° Congresso AITLA*, Guerra, Perugia, pp. 107-126.

Giannoni D.S. 2014, *A comparison of British and Italian customer-complaint forms*, in "English for Specific Purposes" 34, pp. 48-57.

Gray R., Javad M., Power D.M. and Sinclair C.D. 2001, *Social and environmental disclosures and corporate characteristics: A research note and extension*, in "Journal of Business, Finance and Accounting" 28 [3/4], pp. 327-356.

Higgins C. and Walker R. 2012, *Ethos, logos, pathos: Strategies of persuasion in social/environmental reports*, in "Accounting Forum" 36 [3], pp. 194-208.

Höllerer M.A., Jancsary D., Meyer R. E. and Vettori O. 2013, *Imageries of corporate social responsibility: Visual recontextualization and field-level meaning*, in "Institutional Logics in Action" 39 [B], pp. 139-174.

Kramsch C. 2003, *Language and Culture*, Oxford University Press, Oxford.

Laviosa S. 2002, *Corpus-based Translation studies: Theory, Findings, Applications*, Rodopi, Amsterdam/Atlanta.

Leech G. 1983, *Principles of Pragmatics*, Longman, London.

Magistro E. 2013, *The challenges of 'translating' polite discourse for the EU multilingual community*, in "International Journal of Applied Linguistics" 23 [1], pp. 60-79.

Maier C.D. 2014, *Multimodal aspects of corporate social responsibility communication*, in "LEA - Lingue e Letterature d'Oriente e d'Occidente" 3, pp. 245-258.

Malavasi D. 2011, *'Doing well by doing good': A comparative analysis of Nokia's and Ericsson's corporate social responsibility reports*, in Garzone G. and Gotti M. (eds.), *Discourse, Communication and the Enterprise. Genres and Trends*, Peter Lang, Bern, pp. 193-212.

Malavasi D. 2012, *The necessary balance between sustainability and economic success: an analysis of Fiat's and Toyota's corporate social responsibility reports,* in Heynderickx P., Dieltjens S., Jacobs G., Gillaerts P. and de Groot, E. (eds.), *The Language Factor in International Business: New Perspectives on Research, Teaching and Practice*, Peter Lang, Bern, pp. 247-264.

Malavasi D. 2017, *Building trust through corporate identity: An analysis of CSR reports and webpages*, in Salvi R. and Turnbull J. (eds.), *The Discursive Construal of Trust in the Dynamics of Knowledge Diffusion*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 131-153.

Mauranen A. 2004, *Corpora, Universals and Interference*, in Mauranen A. and Kujamäki P. (eds.), *Translation Universals. Do They Exist?*, John Benjamins, Amsterdam/Philadelphia, pp. 65-82.

Newmark P. 1998, *A Textbook of Translation*, Prentice Hall, Hertfordshire.

Ocler R. 2009, *Discourse analysis and corporate social responsibility: a qualitative approach*, in "Society and Business Review" 4 [3], pp.175-186.

Rayson P. and Garside R. 2000, *Comparing corpora using frequency profiling*, in Proceedings of the workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000). 1-8 October 2000, Hong Kong, pp. 1-6.

Skulstad A.S. 2008, *Creating a 'green' image in the public sphere: Corporate environmental reports in a genre perspective*, in Wodak R. and Koller V. (eds.), *Handbook of Communication in the Public Sphere*, De Gruyter Mouton, Berlin, pp. 181-201.

Toury G. 1995, *Descriptive Translation Studies and Beyond*, John Benjamins, Amsterdam/Philadelphia.

Van Leeuwen T. 1996, *The representation of social actors*, in Caldas-Coulthard C.R. and Coulthard M. (eds.), *Texts and Practices: Readings in Critical Discourse Analysis*, Routledge, London, pp. 32-70.

Van Leeuwen T. 2007, *Legitimation in discourse and communication*, in "Discourse and Communication" 1 [1], pp. 91-112.

Venuti L. 1995, *The Translator's Invisibility: A History of Translation*, Routledge, London.

Vergaro C. 2002, *"Dear Sirs, what would you do if you were in our position?" Discourse strategies in Italian and English money-chasing letters*, in "Journal of Pragmatics" 34, pp. 1211-1233.

Vergaro C. 2004, *Discourse strategies in Italian and English sales promotion letters*, in "English for Specific Purposes" 23, pp. 181-207.

Vergaro C. 2005, *Dear sirs, I hope you will find this information useful: discourse strategies in Italian and English "For Your Information" (FYI) letters*, in "Discourse Studies" 7 [1], pp. 119-135.

Wodak R. 2012, *Language, power and identity*, in "Language Teaching" 45 [2], pp. 215-33.

Wodak R., De Cillia R., Reisigl M., Liebhart K. and Hirsch A. 2009, *The Discursive Construction of National Identity* (2nd edition), Edinburgh University Press, Edinburgh.

Wodak R. and Weiss G. 2005, *Analysing European Union discourses: theories and applications*, in Wodak R. and Chilton P. (eds.), *A new agenda in (critical) discourse analysis*, John Benjamins, Amsterdam/Philadelphia, pp. 121-135.

Yu D. and Bondi M. 2017, *The generic structure of CSR reports in Italian, Chinese, and English: A corpus-based analysis*, in "IEEE Transactions on Professional Communication" 60, pp. 273-291.

Zanettin F. 2012, *Translation-driven Corpora*, St. Jerome Publishing, Manchester.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 225-248
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p225

ENGLISH VERSIONS OF CORPORATE WEBSITES
A linguacultural contrastive study of Germany and Spain

COSTANZA CUCCHI
UNIVERSITÀ CATTOLICA DEL SACRO CUORE

Abstract – If local companies in Europe target international visitors as well as national ones, they generally set up an English version of their corporate websites alongside the version in their native language. In their attempt to portray themselves in ways that are engaging to visitors, companies address topics such as corporate history, values and practices on their websites. However, the relevance given to the various topics and the style used may vary across cultures. The present study aims, first, to unveil whether discursive differences exist in website versions in English which were set up by European local companies. Second, it is investigated whether the differences may be explained with reference to Hofstede's model. Because of its importance in Europe, the companies belong to the dairy sector. The countries chosen for the present investigation are Germany and Spain, the cultural differences of which, with reference to Hofstede's model (Hofstede *et al*. 2010), are less marked when contrasted with the countries in previous linguistic studies on corporate websites (Cucchi 2010a, 2012, 2016, 2019). Methodologically, the study, based on a corpus of self-representative discourse from the websites of 12 German and 12 Spanish companies, draws on the Corpus-Assisted Discourse Studies tradition (Partington 2004; Baker 2006), so as to verify the extent to which the findings are compatible with those obtained, within linguistics, from other countries and/or other genres, and with those in other disciplinary domains. The study relies on *Wmatrix* (Rayson 2009) for content analysis and on *WordSmith Tools* (Scott 2012) for the analysis of dispersion plots. Results show that German websites are more informal and give more prominence to data, while Spanish websites rely more on 'self-celebratory' discourse, emphasising tradition, quality and awards. Overall, the findings show that Hofstede's model is helpful for the study of the English versions of websites of European local companies, even in countries where cultural differences are less marked. Since cross-cultural differences still exist among the websites of European companies, future research should address the issue of what content and style are appropriate when targeting international customers in English used as a lingua franca.

Keywords: corporate websites; Hofstede's model; ELF; BELF; language and culture.

1. Introduction

When companies target foreign customers, they need to solve the “dilemma” (de Mooij 2010, p. 14) regarding whether to standardise one or more elements of the marketing mix, e.g. the product, advertising motives and advertising style, or to adapt them to the target culture(s). Standardisation will obviously cut costs and help maintain a unified brand image, but adaptation will arguably attract more customers.

The Standardisation-Adaptation dilemma clearly affects the translation of advertising. De Mooij (e.g. 2004; 2010), who has written extensively on the influence of culture on marketing, argues that advertising should be adapted to be fully effective. In particular, the adaptation process may involve substantial changes in advertising texts and, therefore, translators should cooperate with copywriters in a joint, interdisciplinary effort and be able to “advise about culture-specific aspects” (de Mooij 2004, p. 179). These aspects are advertising appeals – “the values and motives that define the central message” (de Mooij 2010, p. 217) – and communication styles and, according to de Mooij (e.g. 2004, pp. 184, 193; 2010, pp. 221-239, pp. 165-173), their differences across cultures may be understood with reference to Hofstede’s model.

The Standardisation-Adaptation dilemma also affects corporate websites. As Pym (2010, p.1) observed, website localisation, which involves translating and adapting content, is also an interdisciplinary field, which “often involve[s] marketing, design, and software engineering, as well as linguistic processes”. The views on the elements which need localising vary greatly. For example, Pym mentions date and currency formats, colour and images, legal conventions and the insertion of local content, such as “news on the opening of new offices in Paris only for the French version of the site” (Pym 2010, p. 4). In their work, in Pym’s words (2010, p. 12) one of “[t]he few textbooks that address website localization”, Singh and Pereira (2005) offer a much more radical view on localisation and provide guidelines regarding website features (e.g. clubs and chatrooms, free trials) and topics (e.g. family and awards) to be included in websites targeting consumers in specific countries. The guidelines are based on the *Cultural Value Framework* devised by the two authors, which draws on Hofstede’s (Hofstede 2001) and Hall’s (Hall and Reed Hall 1990) cultural models.

Beside being used in marketing and in website localisation, Hofstede’s model was also used for research in linguistics and communications studies (e.g. Bjørge 2007; Crawford Camiciottoli 2017; Dekker *et al.* 2008; Gudykunst and Ting Toomey 1988; Hooker 2008; Kang and Mastin 2008; Katan 2004, 2006; Loukianenko Wolfe 2008; Manca 2016a, 2016b; Vishwanath 2003). However, a limited amount of cross-referencing was noted both among linguistics and communication studies and between them and advertising and localisation studies. If this, on the one hand, enables one to verify the extent to which similar hypotheses and observations were made independently by different authors, on the other hand, there is a need to verify whether the findings from different disciplinary fields are consistent.

The present investigation intends to explore, first, the extent to which the English website versions of selected European companies exhibit different discursive choices, both in content and in style and, second, whether Hofstede's model may account for the differences. In particular, the study investigates website discourse of companies in Germany and Spain. The two countries were chosen so as to verify whether previous findings on corporate websites in Europe – based on Italy and the U.K. (Cucchi 2010a), Greece and Sweden (Cucchi 2012), Austria, the Netherlands, Poland and Portugal (Cucchi 2016), Austria, Denmark, Poland and Portugal (Cucchi 2019),– may be extended to countries which are less far apart along Hofstede's dimensions.

Methodologically, the present study aims to verify whether the findings which emerge using Corpus Assisted Discourse Studies (Partington 2004; Baker 2006), which is a bottom-up approach rooted in linguistic data, are consistent with those which emerged from previous studies relying on different methodologies – a case study on a single website (Cucchi 2010a), a top-down corpus study aimed to verify pre-formulated hypotheses (Cucchi 2012), a corpus investigation of a single noun phrase (Cucchi 2016) and a qualitative investigation of corporate homepages (Cucchi 2019). Clearly, if similar findings result from different data sets and with different methodologies, the findings may be considered reliable. A further aim of the study is to contrast the findings with those from selected studies in the fields of linguistics, communications, localisation and advertising.

2. Review of the literature

Devised in the 1970s to describe differences across cultures, Hofstede's model is characterised by scores along cultural dimensions, which makes cross-cultural comparisons particularly convenient. The model originally consisted of four cultural dimensions, to which further two were later added. The dimensions represent the shared problems that people face, no matter their nationality, but to which they give different responses. The original dimensions are: Individualism (IDV) versus Collectivism (COL) – the relative importance attributed to single individuals or to groups; Power Distance (PD) – the tolerance for power asymmetries; Uncertainty Avoidance (UA) – the tolerance for uncertainty; Masculinity (MAS) versus Femininity (FEM) – the preference for success and tough behaviour versus for the quality of life and modest behavior. The dimensions which were later added to the model are Long Term Orientation (LTO) versus Short Term Orientation (STO) – an orientation towards the future versus an orientation towards the present and the past (Hofstede 2001) – and Indulgence versus Restraint (IVR) – a tendency to enjoy life and have fun versus a belief that this tendency should be regulated by norms (Hofstede *et al.* 2010).

Despite various critiques – for example that the model unduly extends to the whole population the findings obtained from questionnaires to IBM employees, on which it is based, that it assumes that culture is a stable attribute of individuals (Piller 2009), that all the individuals of a specific country share the same culture (McSweeney 2002), and that it is outdated (Jones 2007) – the model was found to be widely used in various disciplinary areas: in global marketing, management and international business, it is "far and away the most prevalent culture framework" (Nakata 2009, p. 248).

Several studies in advertising, localisation, communications and linguistics identified links between Hofstede's dimensions and content and/or style. The studies are below classified on the basis of the relations between Hofstede's dimensions and the three components of the context of situation in Halliday and Hasan's model (1989), namely the field – the topics mentioned in the communicative event; the tenor – the kind of relationship existing between the participants in the communicative event; and the medium, a component of the mode (the role played by language) which refers to the extent to which the text exhibits features which pertain to spoken or written discourse.

Regarding the field, the family theme was associated to low IDV. The *Cultural Value Framework* recommended that "[p]ictures of family, pictures of teams of employees, mention of employee teams and emphasis on team and collective work responsibility [...], and emphasis on customers as a family" (Singh *et al*. 2005, p. 145) should characterise websites targeting low IDV countries. Greek corporate websites in English were indeed characterised by the family theme, often in combination with tradition, thus reflecting Greece's low IDV (Cucchi 2012). An apparently contradictory finding is in Manca (2016b, p. 138), where 'famiglia' is less frequent on the Italian official tourism website as compared to its British counterpart, despite Italy's IDV scores being lower than the U.K.'s. However, as Manca did, the finding may be explained with reference to de Mooij and Hofstede (2010, p. 103), who claimed that pictures of families may either reflect high COL or high IDV, "where people are afraid that family values are disappearing".

The themes of innovation and tradition were linked to low and high UA respectively (de Mooij 2010). More occurrences of *experiment*, *invent* and *try* were indeed found on Swedish corporate websites in English than on their Greek counterparts, thus reflecting Sweden's lower UA compared to Greece's (Cucchi 2012). The *Cultural Value Framework* recommended that themes such as tradition, nostalgia, country history, veneration of the elderly and the culture should be emphasised when targeting high UA countries and phrases like ‘most respected company’, ‘keeping the tradition alive’, ‘for generations’, ‘company legacy’” (Singh *et al.* 2005, p. 102, p. 145) should be used. In keeping with this, Manca (2016b, p. 139) found that *tradizione* and *tradizionale* occurred more frequently on the Italian tourism website than the corresponding words on the British and Australian ones, thus mirroring Italy’s higher UA compared to the U.K.’s and Australia’s.

Interestingly, however, the *Cultural Value Framework* attributed motives which may be easily associated with tradition – pictures of national identity, like flags, historic monuments and other images suggesting the “uniqueness of the country” (Singh *et al.* 2005, p. 145; Singh, Pereira 2005, p. 72) – to low IDV rather than to high UA. Similarly, Cucchi (2012) hypothesised that, due to a tendency to stress what is shared, emphasis on tradition may be linked to low IDV. In the same study, more occurrences of *tradition** and *classic** were found on Greek corporate websites in English than in their Swedish counterparts, reflecting Greece’s lower IDV. While in many European countries lower IDV correlates with higher UA (Meeuwesen *et al.*2009), which could explain why the tradition theme was associated to low IDV in some studies and to high UA in others, it should be noted that mentions of tradition and of recipes, which are shared by the members of a specific community, characterise Portuguese corporate homepages contrasted to Polish ones (Cucchi 2019). Since Portugal and Poland have similarly high UA scores but Portugal’s score on IDV is significantly lower than Poland’s, this finding suggests an association between the tradition theme and low IDV.

The mention of prizes and of quality were variously related to high PD, high UA, high MAS and high IDV in the literature. The *Cultural Value Framework* associated the “[m]ention of awards won, [...] of quality assurance information and quality certification by international and local agencies” to high PD (Singh *et al.* 2005, p. 146; Singh, Pereira 2005, p. 112), since these were interpreted as satisfying the “higher expectation for reassurance” which is considered typical of high PD countries (Singh *et al.* 2005: 130). However, on the basis of its very definition, it would seem more straightforward to associate a need for reassurance to high UA. Drawing on Hofstede (2001, p. 315; 2003, p. 79), a third association can be made, namely between prizes and high MAS, since prizes can be a means to highlight one’s best qualities. Although she did not mention prizes, de Mooij (2010, p. 227) stated that, in advertising, high MAS, when associated with high IDV – “the configuration of Anglo-German cultures” – “leads to a strong need to win, to be successful and show it, combined with the need to dominate”. Cucchi (2019) found that prizes and certificates were significantly present on Polish homepages in English, a finding which is compatible with Poland’s high scores on PD, UA, MAS and IDV. Instead, Cucchi (2012) revealed that *quality*, *award**, *certificate**, *medal** and *prize** were more frequent on Greek corporate websites in English than in their Swedish equivalents. This suggests a possible association between prizes and quality, on the one hand, and high PD, high UA and high MAS, which characterise Greece, but not high IDV, since this is low in Greece. However, the significant presence of prizes and certificates found on Portuguese homepages in English (Cucchi 2019) suggests a link between prizes and certificates, high UA and high PD, since Portugal scores high on UA and PD, but low on MAS and very low on IDV.

According to de Mooij (2010, p. 233), purity and freshness are popular advertising appeals for the food products of high UA countries, where people generally have a more passive attitude to health than people in low UA countries, who focus more on fitness and sport (de Mooij, Hofstede 2010, p. 90). Interestingly, purity and freshness may reassure customers about the quality of the products, which would strengthen the hypothesis of an association between quality assurance and prizes and high UA. In keeping with this interpretation, Manca (2016b, p. 140) associated the higher frequency of *natura* on the Italian tourism website compared to *nature* on its British and Australian equivalents to a greater need for purity typical of Italy, in keeping with its higher UA.

A need for safety was related to high UA. In high UA countries, advice from experts is valued (Singh, Pereira 2005, p. 98) and, in advertising, manufacturers should prove that they are competent (de Mooij 2010, p. 233). High UA was described as "translat[ing] into the need for explanations, structure, testing, test reports, scientific proof and advice, and testimonials by experts, but also into high technology and design, the latter even more in configuration with high power distance" (de Mooij 2010, p. 232). Various studies in communications and linguistics supported these claims. High UA was found to be linked to the amount of details considered as desirable in on-line eBay auctions (Vishwanath 2003). As compared to British brochures for pension funds, more dependency on expertise and need for advice was found in Italian brochures, thus reflecting Italy's higher UA (Katan 2006). However, it should be noted that the amount of details may not only be associated to higher UA, but also to higher PD, since the latter arguably leads to a greater need by companies to portray themselves as experts (Cucchi 2010a). In this respect, it is worth mentioning that, in European countries, higher UA and higher PD tend to correlate (Meeuwesen *et al.* 2009).

LTO and IVR were not operationalised in the *Cultural Value Framework*. While IVR is still young, having been introduced only in 2010, LTO seems difficult to understand and operationalise. Hofstede (2001, p. 355) himself claimed that this dimension, which refers to typically Eastern values, "will probably puzzle many Western readers". Despite the association of LTO to future rewards and of STO to the past and the present made in Hofstede (2001), de Mooij (2010) discarded the past component inherent in STO, linking it to "the sense of urgency so often encountered in U.S. advertising", reflected in expressions such as 'Hurry', 'Don't wait', and the emphasis on "living in the now and not thinking about the future" (de Mooij 2010, p. 236).

Linguistics and communications studies unanimously related lower PD to more interacting roles in terms of tenor and to a more marked spoken medium in terms of mode. Reportedly, people in low PD societies expect "more responsiveness and empathy" (Singh *et al.* 2005, p. 130). In keeping with this, in lower PD countries roles in doctor-patient consultations were less fixed (Meuweesen *et al.* 2009) and virtual team members were more responsive to messages and deemed it more important to invite contributions to the group (Dekker *et al.* 2008). In American sales promotion letters compared to Russian ones, more attempts to "continue the conversation" were observed, in keeping with the U.S.'s lower PD compared to Russia's (Loukianenko Wolfe 2008). The official tourism homepages of New Zealand, Germany, France and Italy reflected their PD scores, with New Zealand's homepage being by far the most interactive one (Manca 2016a).

In lower PD countries a more conversational medium was also found. More informal greetings and closings were observed in e-mails by students from lower PD countries to their university professors (Bjørge 2007) and more occurrences of the vague markers *and so on* and *etcetera* – which are typical of spoken language (Carter, McCarthy 2006, p. 202) – were found in EU parliamentary discourse by British speakers compared to their Italian equivalents *e così via* and *eccetera* in Italian EU parliamentary discourse, which reflected Italy's higher PD (Cucchi 2010b). Tourism public relation websites of low PD countries featured more casual narratives (Kang, Mastin 2008), which also characterised a British website compared to the English version of its Italian counterpart (Cucchi 2010a). More occurrences of *thing(s)* – one of the most frequent words in spoken English (Carter, McCarthy 1997, p. 16) – were found on Swedish corporate websites in English as compared to Greek ones, which reflected Sweden's much lower PD. Corporate homepages in English of Austrian, Danish, Polish and Portuguese companies reflected the PD of the four countries, with Austria's homepages containing many features of spokenness, such as second person pronouns, imperatives, questions, vague language, exclamation marks and suspension marks (Cucchi 2019).

Drawing on Gudykunst and Ting Toomey (1988), de Mooij (2010, pp. 165-166) associated specific styles, as she called them, to a combination of dimensions. High IDV, low PD and low-context communication, in which messages tend to be communicated explicitly (Hall 1976; Hall, Reed Hall 1990), was associated to the *verbal textual style*, which focuses on "personhood", consists in "individual-centred language, [...] [which] enhances the 'I' identity and is person-oriented (e.g. English)" (de Mooij 2010, p. 165). Low IDV, high PD and high-context, more implicit communication was linked to the *verbal contextual style*, in which the language is “role-centered [...]; [...] emphasises a context-related role identity (e.g. Japanese, Chinese), [...] includes different ways of addressing different persons, related to their status and/or situation” (de Mooij 2010, p. 165).

A further distinction was claimed to exist between the *exacting style*, the *elaborate style* and the *succinct style*. The *exacting style*, also dubbed *precise*, characterises low PD countries, namely Anglo-American and Northern European ones: people “like data” (de Mooij 2010, p. 167), communication is direct and “no more or no less information than required” is provided (de Mooij 2010, p. 165), a definition which appears culture-bound – required by whom? – since the Author is Dutch. On the contrary, in low IDV countries communication is more indirect and implicit. In particular, low IDV countries with high PD and high UA, like Japan, Spain, Italy and the Arab countries, generally utilise the *elaborate style*, which is flowery, characterised by metaphors, long strings of adjectives and proverbs. Instead, low IDV countries with high PD and low UA, like India, China and Singapore, typically use the *succinct style*, characterised by understatements, pauses and silence (de Mooij 2010, p. 167).

Another “major distinction” made by de Mooij (2010, p. 170) was between the *direct style* of high IDV countries and the *indirect style* of low IDV countries. In high IDV countries, “the public tends to be addressed in a direct and personalized way. Words like *you*, *we* and *I* are frequently used. So are imperatives” (de Mooij 2010, p. 223). In low IDV cultures the public is more frequently addressed through indirect methods, such as drama or metaphors (de Mooij 2010, p. 170). The association between high IDV and a more personalised style also emerged from various communications and linguistics studies. More occurrences of *I*, *my*, *you* and *your* were observed on U.S. websites as compared to Chinese ones, in keeping with the U.S.’s higher IDV (Zhao *et al.*2003). More occurrences of *you*, *your*, *we*, *us*, *our* were found on a British corporate website as compared to its Italian counterpart written in English (Cucchi 2010a) and on Swedish corporate websites in English compared to their Greek counterparts (Cucchi 2012), which reflected the U.K.’s and Sweden’s lower PD compared to Italy’s and Greece’s, respectively. *Welcome*, *contact* and *call* to address customers were also more frequent on Swedish websites in English than in Greek ones (Cucchi 2012). In keeping with these findings, Manca (2016b, p. 103) found that imperatives were absent on the Italian tourist website, but played a significant role in its British and Australian equivalents. It is here worth mentioning that not only are IDV scores in Australia, the U.K. and Sweden higher than in Italy and Greece, but the former three countries’ PD and UA are also lower than in Italy and in Greece. Interestingly, the predictions which can be formulated on the basis of these three dimensions are the same. High IDV was associated, as seen above, to more personalisation. Lower PD may be linked to more interactive roles and, therefore, to more imperatives and personal pronouns used to interact with website visitors. In turn, more interactive roles may be linked to lower UA, due to a greater tolerance for unpredictable situations resulting from increased interaction.

De Mooij (2010, p. 167) associated, respectively, "[w]ordings such as *absolutely* and *definitely* to express buying intentions" with the direct style and high IDV, and *probably* and *somewhat* with the indirect style and low IDV. Drawing on this claim, Crawford Camiciottoli (2017) hypothesised that boosters would be more frequent in the earnings conference calls from European companies, while diminishers would be more frequent in Asian companies. In fact, results contradicted this hypothesis, showing that both boosters and diminishers were more frequent in the discourse of European companies. However, it should be noted that boosters and diminishers are both markers of personalised discourse – more informal and conversational than impersonal discourse – which de Mooij associated to high IDV, as seen above. Indeed, Chafe (1982, p. 47) mentioned *really* and *just* – a booster and a hedge – as examples of the 'emphatic particles' which "express enthusiastic involvement in what is being said" and characterise spoken discourse. Similarly, Biber (1988, p. 108) found emphatics, amplifiers and hedges to be typical of spoken texts, characterised by an interactive and involved purpose. Following this interpretation, a higher frequency of both boosters and diminishers would be expected in high IDV countries, where communication is typically more personalised, which is what Crawford Camiciottoli's findings indicate (2007, p. 117).

3. Methodological issues and corpus

To verify whether Hofstede's model may account for the discursive choices made on the English website versions of local companies in Europe, a corpus of German and Spanish corporate websites was set up. The corpus was examined through a CADS approach so as to unveil whether the findings are compatible with those obtained with other methodologies and in other research fields. Before illustrating the rationale for corpus building, it should be mentioned that English as used on local corporate websites may be qualified as ELF (English as a Lingua Franca) or BELF (Business English as a Lingua Franca). Big multinationals generally set up website versions in many different languages to address international customers in their native languages: English is just one of the many available languages. Some multinationals (e.g. Nestlé) even exhibit various versions in English which target consumers from different geographic areas, for example, the U.K., the U.S. and India. Instead, local companies typically set up a single translation into English of their websites along with the original version in their national language.

The purpose of the English version is clearly to target consumers whose mother tongue is different from the national language of the country in which the companies operate. Therefore, English functions as a lingua franca, intended as "the common language" which enables communication "among speakers who come from different linguacultural backgrounds" (Jenkins 2009, p. 200). Given that the specific function of English on national corporate websites consists in "conducting business within the global business discourse community" (Louhiala-Salminen *et al*. 2005, p. 404), English may be labelled as BELF (Business English as a Lingua Franca), a term coined as a result of the transfer of the ELF paradigm to the business context (Louhiala-Salminen *et al.* 2005). Unlike what happens on the websites of multinationals, on the websites of local companies English is used to reach consumers who belong, in Kachru's (1992) terminology, to the Inner Circle, the Outer Circle and the Expanding Circle, i.e. respectively, native, second language, and foreign language speakers of English. This is in line with current conceptualisations of ELF (e.g. Hülmbauer *et al.* 2008, Seidlhofer 2009, Jenkins 2009; European Commission 2011), which promote an inclusive view of ELF, intended as a means of communication between speakers of all the three Circles. In terms of form, the ELF and BELF paradigm is characterised by emphasis on effectiveness rather than on native-like correctness (e.g. Kankaaranta and Planken 2010, European Commission 2011), which is reflected in the presence of minor grammatical and lexical deviations on corporate websites.

Germany and Spain were chosen for analysis because the two countries – located in Europe, a geographically limited territory in which cultural differences may be thought to be blurred due to globalisation – exhibit the correlation among the dimensions which is typical of European countries, with higher IDV, lower PD and lower UA and the other way round (Meeuwesen *et al.* 2009). In addition, the two countries differ along Hofstede's dimensions (Table 1), but not so markedly as the countries considered in previous studies on English website versions in Europe (Cucchi 2010a, 2012, 2016, 2019).

Countries	*IDV*	*PD*	*UA*	*MAS*	*LTO*	*IRV*
Germany	67	35	65	66	83	40
Spain	51	57	86	42	48	44

Table 1
Country scores.[1]

[1] The scores are taken from Hofstede *et al.*'s (2010), which offers comprehensive grids for the scores on all the dimensions but LTO. For this dimensions, scores are reported (p. 240) for only 23 countries and Spain is not among these. LTO scores for Spain are obtained from *Hofstede*

For an interpretation of the scores, it should be mentioned that countries are measured on a scale from 1 to 100 and that a few scores in the model exceed 100, since the respective countries were added later. Scores below 50 are conventionally considered low, those above 50 are considered high. To fully appreciate the scores, the highest and lowest scores both in Europe and in the world are worth mentioning. U.K.'s IDV (89) and the U.S.'s. (91) are the highest in Europe and in the world, respectively, while Serbia (IDV=25) and Guatemala (IDV=6) score the lowest. Austria's PD (11) is the lowest in the world, while Slovakia's and Malaysia's (104) is the highest. The highest UA in the world is Greece's (112), while Denmark (UA=23) and Singapore (UA=8) score lowest, in Europe and in the world respectively. Slovakia is the most MAS country in the world (MAS=110) and Sweden the least one (MAS=5). Sweden (78) is the most indulgent country in Europe and Venezuela (IVR=100) in the world, while Latvia (IVR=13) and Pakistan (IVR=0) are the least indulgent countries in Europe and in the world.[2]

The cheese sector was chosen because of its importance in Europe. According to the *European Food Information Council*, "[c]heese, when consumed in moderation, is an important part of a healthy, balanced diet and at the centre of European culture".[3] Given the difficulties in developing "a workable method to identify representative [...] discourse in an erratic Web site universe" (Isaksson, Jørgensen 2010, p. 128), the national dairy associations mentioned on the website of the *European Dairy Association* were considered representative of the dairy sector in each country.[4]

The member companies of the German and Spanish national dairy associations were examined, for a total of 90 and 47 companies respectively. Companies which did not deal with cheese, multinationals, local companies whose websites did not have an English version and the companies whose website was protected and could not be copied in a word file were eliminated. This yielded 12 German and 12 Spanish companies.

The following texts were included in the corpus:

- Self-representative texts describing values, history, achievements, production, corporate and social responsibility, news;
- Texts regulating the company's relationship with visitors and customers, e.g. sections about data protection and training.

Insights (https://www.hofstede-insights.com/product/compare-countries/), which enables comparison among the dimensions of up to four countries.

[2] Neither Hofstede *et al.* (2010) nor *Hofstede Insights* (https://www.hofstede-insights.com/product/compare-countries/) provide comprehensive grids with LTO scores. Therefore, overall comparisons along this dimension are not included in this study.

[3] https://www.eufic.org/en/food-today/article/cheese-a-european-tradition

[4] http://www.euromilk.org/eda/members.aspx?cid=15

To ensure greater comparability, product descriptions were not included. The corpus is illustrated in Table 2 and consists of two sections, which contain respectively the English versions of German and Spanish websites. The two sections were called DE and ES with reference to the two-letter country code used to identify the Member States of the European Union (*Eurostat*, Statistics Explained).

Although the number of websites included in the two sections is equal, the number of tokens is much lower in ES, which indicates that less information is generally displayed on Spanish websites. This is in keeping with the fact that Spain is, compared to Germany, a high-context culture: more information is conveyed verbally or not conveyed at all, since it is considered shared.

DE *35.308 tokens*		*ES* *17.121 tokens*	
Alpenhein	Exquisa	Aldonza y DonIsmael	Los Cameros
Ammerland	Goldsteig	Coqueya	Maxorata
Andechser Natur	Kohrener	El Gran Cardenal	Monteverdesa
Bauer Milch	Petrella	El Pastor	Palancares
Bechtel	Rücker	Eusebio Manzano	Pastoret
Bergader	Weideglück	Garcia Baquero	San Vincente

Table 2
The corpus.

Content analysis was conducted with *Wmatrix* (Rayson 2009), since this software enables identification not only of the most frequent words, as *WordSmith Tools* (Scott 2012) does, but also of the most frequent semantic domains and of the words belonging to them, which makes content analysis more accurate. As a part of content analysis, dispersion plots were examined with *WordSmith Tools* (Scott 2012). Compared to *Wmatrix*, which can only process the two sections of the corpora as two individual files, *WordSmith Tools* makes it possible to keep the website texts of the single companies separate. Dispersion plots, one of the facilities of *WordSmith Tools*, reveal the distribution of the investigated features across the corpus: the significance of a given feature is clearly very different if it occurs across a large number of companies or, albeit very frequently, in very few of them. In Section 4, data about the dispersion plots is given in the form of the number of companies where specific features appear. Section 4 compares the findings from content analysis with those emerged from previous studies – from various disciplinary fields – which relied on Hofstede's framework, to find out whether content and stylistic features may be explained with reference to the framework.

4. Content Analysis and Discussion

4.1. Semantic Domains

Table 3 shows the top ten semantic domains along with the top ten words in each domain, in DE and ES.

Rank	*DE semantic domains*	*DE top ten words*	*ES semantic domains*	*ES top ten words*
1	Grammatical bin 28.84%	the, and, of, to, in, for, a, with, is, or	Grammatical bin 30.84%	the, of, and, in, to, a, with, for, from
2	Pronouns 5.75%	our, we, you, your, that, it, us, this, its, which	Pronouns 5.39%	our, its, it, we, that, which, you, us, their, who
3	Unmatched 4.64%	Bergader, Goldsteig, Alpenhain, Karwendel, Ammerland, Schwarzenfeld, Exquisa, Andechser, Bauer, DLG	Unmatched 4.84%	Pastoret, Cameros, industrias, cheesemaking, manchego, lacteas, Lcteos, lcteas, Oscos, villacenteno
4	Numbers 2.52%	0,+49, 1, one, 3, 2, 6, three, 5	Food 4.08%	cheese, cheeses, food, yougurt, recipe, curd, dessert, raw, whey, youghurt
5	Food 2.03%	cheese, food, dairy products, cookies, cheeses, yoghurt, butter, raw, cookie, fruit	Personal names 2.52%	Garca_Baquero, Los_Cameros, La_Rioja, del, tel., Los_Oscos, Hijos_de_Salvador, Castilla_La_Mancha, Villanueva_de, El_Gran_Cardenal
6	General actions, making 1.95%	production, processes, process, processing, made, make, activities, committed, waging, making	Numbers 2.24%	one, three, +34, 2010, 2012, 2007,2005, 2008, 1, 4
7	Existing 1.56%	is, are, be, ‘s, available, event, were, been, ’re, was	General actions, making 2.22%	production, process, made, make, processes, manufacturing, makes, making, activities, committed
8	Objects generally 1.41%	products, product, goods, things, packaging, files, mould, trail, provisions, equipment	Geographical names 1.59%	Spanish, S.A., San Vicente, Spain, ewes, la, Greek, ewe, Spains, Canary Island
9	Location and direction 1.35%	this, here, where, top, external, there, direct, northern, located, route	Existing 1.45%	is, are, be, was, were, real, been, available, event, being
10	Personal names 1.32%	tel., Privatmolkerei_Bechtel, Andechser_Molkerei_Scheitz, GOLDSTEIG_Ksereien_Bayerwald, Google_Analytics, Kohrener_Landmolkerei_GmbH, Privatmolkerei_Bauer_GmbH, Molkerei_Ammerland, Hans, european_union	Objects generally 1.35%	products, product, medal, packaging, pack, moulds, packs, wedges, SIAL,[5] medals

Table 3
Top ten semantic domains and top ten corresponding words.

[5] SIAL is an international exhibition in the food sector.

In both DE and ES, pronouns play an important role, being the second semantic domain after the ‘Grammatical bin’, which contains function words (Table 3). However, the personal pronouns and possessive adjectives used to interact with website visitors are invariably more prominent in DE than in ES, both in terms of overall frequency and of the number of companies which use them (Table 4), making DE discourse more informal (Example 1). The higher personalisation in DE may be associated to the *verbal textual style* (de Mooij 2010), which reflects Germany’s higher IDV and lower PD compared to Spain.

Pronouns / possessive adjectives	*DE*	*ES*
our	1.31% all companies	0.92% 11 companies
we	0.92% all companies	0.44% all companies
you	0.60% 11 companies	0.40% 9 companies
your	0.38% all companies	0.12% 5 companies
Total	3.21%	1.88%

Table 4
Selected pronouns and possessive adjectives.

(1) “Heeeeyyy, heeeeere! We are here!” Questions? Questions! Suggestions? Suggestions! Simply fill out the form below and your request is with us. We are happy to help! (Rcker)

The higher frequency of *we* to refer to companies in DE explains why *Personal names*, which comprises company names, is more frequent in ES (Table 3). As a result, sentences in which companies refer to themselves by their names (Example 2, from ES) are more frequent in ES, making ES discourse more impersonal.

(2) Since 1962, Garca Baquero, originally a family business founded by Hersilio Garca Baquero, has specialised in the production of cheese. (Garca Baquero)

Interestingly, unlike in ES, in DE various nouns denoting people collocate with *our* – *customers*, *employees*, *farmers*, *suppliers* and *partners*, thus showing their contribution to the success of the company. These findings, which confirm Cucchi (2010a, 2016), may be attributed to greater attention to single individuals, possibly due to higher IDV, as well as to lower PD, which is associated to symmetrical relationships within companies, as illustrated in (3).

(3) We value our employees highly […]. The tasks carried out by each and every employee are important, and everyone is responsible for achieving our corporate goals. Moreover, we respect the dignity of each and every employee. We expect all our employees to behave in a cooperative way. […] And every member of our management team practises an “open door” policy. (Ammerland)

The semantic domain *Numbers* is more frequent in DE than in ES (Table 3), reflecting the features of the *exacting style* typical of low PD countries (de Mooij 2010).

(4) Within the beautifully appointed landscape of the Rupertiwinkel (up to 827 m above sea level) and close to the old royal city of Burghausen in south-eastern Bavaria there are around 520 farms [...].

Recipe, possibly associated to low IDV countries (Cucchi 2019), features in ES among the top ten words in *Food* (Table 3). Although the frequency of *recipe** is notably higher in ES, the word features in a limited number of companies in both ES and DE (Table 5). However, the importance of traditional values in ES constantly emerges from the co-text of *recipe** (Example 5). These findings may be related to Spain’s lower IDV compared to Germany.

Word	*DE*	*ES*
recipe*	0.01% 5 companies	0.11% 4 companies

Table 5
*Recipe**.

(5) At Pastoret our goal is to make fresh products for you to enjoy. To do this, we have maintained our original family recipes and traditional artisan production techniques. (Pastoret)

Existing comprises the contracted forms *’s* and *’re* in DE (Table 3), which make DE discourse more informal (Example 6), thus reflecting Germany’s lower PD. Although contracted forms feature only in a few German companies, they notably never occur in ES (Table 6).

Contracted forms	*DE*	*ES*
’s	0.04% 5 companies	/
’re	0.02% 3 companies	/
Total	0.06%	/

Table 6
Contracted forms.

(6) If you’re a cow in Ammerland, you’re certain to be a happy cow. Because you’re surrounded by unspoilt nature, luscious meadows and so much space you can graze outside on our green fields. (Ammerland)

Within *Objects, generally*, the informal noun *thing* (Example 7) appears among the top ten words in DE (Table 3). Table 7 shows that *thing** is more frequent in DE, both in terms of frequency and of number of companies, in keeping with Germany’s lower PD. On the contrary, *medal* and *medals* are among the top ten words in *Objects, generally* in ES, but not in DE (Table 3), where their frequency is much lower (Table 7). This result could point to an association between the mention of quality and prizes, on the one hand, and high PD (Singh *et al.* 2005) and high UA, on the other, rather than to high IDV and high MAS, as claimed by de Mooij (2010). Indeed, Spain’s PD and UA are higher than Germany’s, but its IDV and MAS are lower.

Objects, generally	*DE*	*ES*
thing*	0.05% 7 companies	0.01% 1 company
medal*	0.02% 4 companies	0.13% 5 companies

Table 7
Selected objects: *thing** and *medal**.

(7) The tenet of the entire ANDECHSER NATUR product range is “Leave natural things natural!” (Andechser)

(8) El Gran Cardenals best kept secret has been revealed. This secret closely guarded by our brother, we have shared with expert cheesemakers around the world and have honored us in 2009 and 2010 with the gold medal. (El Cardenal)

4.2. Most frequent content words

Table 8 illustrates the top ten content words in DE and in ES, identified with *Wmatrix*. The words in italics were chosen for analysis, since they were mentioned in previous literature on Hofstede’s dimensions.
Table 9 illustrates the frequencies of the selected words and the number of companies in which they feature.

Rank	*DE*	*ES*
1	milk 0.73%	cheese 2.06%
2	cheese 0.69%	*quality* 0.55%
3	*quality* 0.46%	milk 0.54%
4	products 0.44%	cheeses 0.44%
5	dairy 0.40%	products 0.42%
6	production / *data* 0.36%	production 0.39%
7	company 0.31%	company / process 0.33%
8	product 0.30%	*awards* 0.30%
9	training 0.29%	*traditional* 0.29%
10	website 0.28%	*natural* 0.27%

Table 8
Top ten content words.

Words	*DE*	*ES*
quality	0.46% all companies	0.55% all companies
award*	/	0.38% 7 companies
data	0.36% 10 companies	0.12% 3 companies
tradition*	0.09% 8 companies	0.71% 11 companies
natural	0.13% 12 companies	0.27% 9 companies

Table 9
Selected words.

The fact that *quality* is slightly more frequent in ES and that *award** is absent in DE suggests an association between the mention of quality and awards, on the one hand, and higher PD and UA, on the other, as stated in Section 4.1. Examples (9) and (10) illustrate *quality* and *award** in ES.

(9) From the outset the basic philosophy was always, above all, to offer the highest quality product [...]. (Monteverdesa)

(10) Throughout its career, the company and its brands have been awarded numerous prizes and mentions. (Maxorata)

Data, which is more frequent in DE, is embedded in legal discourse regulating the disclosure of personal data (Example 11).

(11) We collect and use your personal data solely in accordance with the provisions of the data protection law of the Federal Republic of Germany. (Bauer)

The greater presence of legal discourse in DE is in keeping with the importance of contracts and explicit rules in *rule-based cultures* (Hooker

2008), which are characterised by lower PD and where human beings are conceived as autonomous individuals. Consequently, no individual has natural authority over others and sources of authority apart from individuals need to be established. Instead, in *relationship-based cultures*, characterised by higher PD, "[i]t is not necessary to write everything (or perhaps anything) down, because mutual understanding and a handshake suffice" (Hooker 2008: 4).

*Tradition** is more frequent in ES and tends to co-occur with other words associated to tradition, like *artisan*, *family*, *original* and *recipe* (Example 12). Instead, in DE *traditional* invariably occurs in combination with words indicating modernity (Example 13). These findings could be related both to Spain's lower IDV and to its higher UA, as in previous literature (Singh *et al.* 2005; Cucchi 2012).

(12) At Pastoret […] we have maintained our original family recipes and *traditional* artisan production techniques. (Pastoret)

(13) Even though we have one of Europe 's largest and most advanced cheese dairies and an ultra-modern plant for fresh dairy products, we still produce by *traditional* methods without exception. (Ammerland)

Natural is more frequent in ES, in keeping with Spain's higher UA, confirming the association made in de Mooij (2010). Example 14 shows *natural* in ES.

(14) Over 35 years ago our cheese began to be produced in the same original way it was made at the very sheepfolds: natural milk and rennet. (Quesos Aldonza y Don Ismael)

5. Concluding remarks

Results show that, despite globalisation, linguacultural differences are still visible in the English versions of corporate websites of two European countries which are not exceedingly far apart on Hofstede's cultural dimensions. In particular, a CADS approach which combined the study of semantic domains, frequent words, dispersion plots and collocations has proved effective for the identification of linguacultural differences which can be explained with reference to Hofstede's framework.

In some cases, a single dimension can explain the field of discourse quite well, for example *natural* was more frequent in ES, reflecting the association between freshness and high UA (de Mooij 2010). In other cases, more dimensions may be associated to specific linguistic choices. For example, the emphasis on tradition in ES may be interpreted as a means of reducing uncertainty, but also of emphasising shared values, and thus be related to high UA (Singh *et al.* 2005, Manca 2016b) and to low IDV (Cucchi 2012, 2019). Similarly, being a sign of authority, the mention of quality and awards in ES may be associated to high PD (Singh *et al.* 2005; Singh, Pereira 2005) and, since it may reduce the customers' uncertainty about the products, to high UA (Cucchi 2012, 2019).

Regarding the tenor and the medium, findings confirm a link between interactive roles, spokenness and low PD. Interestingly, linguistics studies grouped and, at times, even coined specific terms to refer to features co-occurring in spoken discourse, which were found to characterise DE in the present investigation. Chafe (1982, p. 46) considered first person references, including *we*, as a "symptom of involvement", which points to the typical interactive relationship in speech as opposed to writing, characterised by detachment. Vagueness and hedges – a category to which the vague noun *thing* belongs – are also a sign of involvement typical of speech (Chafe 1982, p. 48). Biber (1988; p. 107) found that first person pronouns, contractions and hedges typically co-occur in spoken texts, which have an "interactive, affective, and involved" purpose. Hyland (2004, p. 113) termed 'engagement markers' the second person pronoun *you*, imperatives and questions and defined them as devices that "explicitly address readers, either to focus their attention or to include them as discourse participants". On the basis of these observations, in texts targeting idealised readers like websites and brochures an association may be made between lower PD and *synthetic personalisation*, defined as the practice, popular in advertising, of giving "the impression of treating each of the people 'handled' *en masse* as an individual" (Fairclough 1989, p. 62).

Since English versions of European corporate websites do differ and Hofstede's model has proved to be helpful for explaining the differences, future research could consider whether the versions in English, which reflect national source texts in content and style, are effective in reaching a culturally diverse audience. Clearly, this issue has far-reaching implications for the content and style which is deemed appropriate for an international public and for all those involved in website localisation, including translators.

Bionote: Costanza Cucchi, PhD in Applied Linguistics and Communication, is an Associate Professor in English Language and Translation at the Faculty of Linguistic Sciences and Foreign Literatures at the Catholic University of Milan. She teaches undergraduate courses in English Linguistics, Professional Communication in English and Negotiation Strategies and holds workshops in intercultural communication. Her current research interests are the use of BELF (Business English as a Lingua Franca) in corporate

settings and the relationship between language and identity. She has published and presented papers at national and international conferences on non-native English in the European Parliament, vague language in parliamentary discourse, national identity as manifested in corporate websites and in CLIL textbooks, the teaching of English lexis and international classroom exchanges.

Author's address: costanza.cucchi@unicatt.it

References

Baker P. 2006, *Using Corpora in Discourse Analysis*, Continuum, London.

Biber D. 1988, *Variation across speech and writing*, Cambridge University Press, Cambridge.

Bjørge A.K. 2007, *Power distance in English lingua franca email communication*, in "International Journal of Applied Linguistics" 17, pp. 60-80.

Carter R. and McCarthy M. 1997, *Exploring Spoken English*, Cambridge University Press, Cambridge.

Carter R. and McCarthy M. 2006, *Cambridge grammar of English: A comprehensive guide. Spoken and written English grammar and usage*, Cambridge University Press, Cambridge.

Crawford Camiciottoli B. 2017, *The interpersonal pragmatics of intercultural financial discourse.A contrastive analysis of European vs. Asian earnings conference calls*, in Kecskes I. and Assimakopoulos S. (eds.), *Current Issues in Intercultural Pragmatics*, John Benjamins, Amsterdam/Philadelphia, pp. 105-127.

Chafe W. 1982, *Integration and involvement in Speaking, Writing, and Oral Literature*, Tannen D. (ed.), *Spoken and Written Language: Exploring Orality and Literacy*, Ablex Publishing Corporation, Norwood, New Jersey, pp. 35-53.

Cucchi C. 2010a, *Hofstede's cultural dimensions: Italian national identity in ELF usage*, in "Cultus. The Journal of Intercultural Mediation and Communication" 3, pp. 137-158. http://www.cultusjournal.com/files/Archives/costanza_cucchi.pdf (15.10.2018).

Cucchi C. 2010b, *Vague expressions in the European Parliament: A marker of cultural identity?*, in Garzone G. and Archibald J. (eds.), *Discourse, identities and roles in specialized communication*, Peter Lang, Bern, pp. 85-107.

Cucchi C. 2012, *Hofstede's cultural dimensions in ELF company websites of European countries: lexical choices in Sweden and Greece*, in Facchinetti R. (ed.), *A Cultural Journey through the English Lexicon*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 153-181.

Cucchi C. 2016, *BELF and national cultures on European corporate websites: A cross-cultural investigation*, in Lopriore L. and Grazzi E. (eds.), *Intercultural Communication. New Perspectives from ELF*, Roma Tre-Press, Roma, pp. 385-497. http://romatrepress.uniroma3.it/ojs/index.php/elf/article/view/602/599 (15.10.2018).

Cucchi C. 2019, *National Cultures on European Corporate Homepages in English: A Linguistic Analysis*, in "International Journal of Business Communication" 56 [2], pp. 198-232.

Dekker D.M., Rutte C.G. and van den Berg P.T. 2008, *Cultural differences in the perception of critical interaction behaviors in global virtual teams*, in "International Journal of Intercultural Relations" 32, pp. 441-452.

de Mooij M. 2004, *Translating Advertising. Painting the Tip of an Iceberg*, in "The Translator" 10 [2], pp. 179-198. https://www.mariekedemooij.com/articles/demooij_2004_translator.pdf (15.10.2018).

de Mooij M. 2010, *Global Marketing and Advertising. Understanding Cultural Paradoxes*, 3rd edition, Sage Publications, Thousands Oaks.

de Mooij M. and Hofstede 2010, *The Hofstede model. Applications to global branding and advertising strategy and research*, in "International Journal of Advertising" 29 [1], pp. 85-110.

European Dairy Association, http://www.euromilk.org/eda/index.aspx (15.10.2018).

European Food Information Council, 12 October 2009, *Cheese: a European tradition*. https://www.eufic.org/en/food-today/article/cheese-a-european-tradition (15.10.2018).

European Commission, Directorate-General for Translation 2011, *Lingua Franca: Chimera or Reality?* http://translationjournal.net/images/e-Books/PDF_Files/Lingua%20Franca%20Chimera%20or%20Reality.pdf (15.10.2018).

Eurostat, Statistics Explained, 19 February 2018, *Glossary: Country codes*. https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Country_codes (15.10.2018).

Gudykunst W. and Ting Toomey S. 1988, *Culture and Interpersonal Communication*, Sage Publications, Newbury Park, CA.

Fairclough N. 1989, *Language and Power*, Longman, New York.

Hall E.T. 1977, *Beyond Culture*, Anchor Press/Doubleday, Garden City, NJ.

Hall E.T. and Reed Hall M. 1990, *Understanding cultural differences: Germans, French and Americans*, Intercultural Press, Boston, MA.

Halliday, M.A.K. and Hasan R. 1989, *Language, context and text: Aspects of language in a social-semiotic perspective*, Oxford University Press, Oxford.

Hofstede G. 2001, *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*, Sage, Thousand Oaks.

Hofstede G. 2003, *Cultures and organizations: Software of the mind, intercultural cooperation and its importance for survival*, Profile Books, London. (Original work published 1991).

Hofstede G., Hofstede G. J. and Minkov M. 2010, *Cultures and organizations: Software of the mind, intercultural cooperation and its importance for survival*, 3rd edition, McGraw-Hill, New York.

Hofstede Insights, https://www.hofstede-insights.com/ (15.10.2018).

Hooker J. 2008, *Cultural Differences in Business Communication*, pp. 1-19. http://public.tepper.cmu.edu/jnh/businessCommunication.pdf (15.10.2018).

Hülmbauer C., Böhringer H. and Seidlhofer B. 2008, *Introducing English as a lingua franca (ELF): Precursor and partner in intercultural communication*, in. "Synergies Europe" 3, pp. 25-36. https://gerflint.fr/Base/Europe3/hulmbauer.pdf (15.10.2018).

Hyland K. 2004, *Disciplinary discourses: Social interactions in academic writing*, University of Michigan Press, Ann Arbor.

Isaksson M. and Jørgensen F.P.E. 2010, *Communicating corporate ethos on the web: The self-presentation of PR agencies*, in "Journal of Business Communication" 47, pp. 119-140.

Jenkins J. 2009, *English as a Lingua Franca: interpretations and attitudes*, in "World Englishes" 28 [2], pp. 200-207.

Jones M.L. 2007, *Hofstede – Culturally questionable?* in "Research Online". http://ro.uow.edu.au/commpapers/370 (15.10.2018). Originally published as Jones, M. 2007, *Hofstede – Culturally questionable?* Oxford Business & Economics Conference. Oxford, UK, 24-26 June, 2007.

Kachru B.B. 1992, *Models for non-native Englishes*, in Kachru B.B. (ed.), *The other tongue: English across cultures* (2nd edition), Urbana and Chicago, University of Illinois Press, pp. 48-74.

Kang D.S. and Mastin T. 2008, *How cultural difference affects international tourism public relations websites: A comparative analysis using Hofstede's cultural dimensions. Short communication*, in "Public Relations Review" 34, pp. 54-56.

Kankaanranta A. and Planken B. 2010, *Belf Competence as Business Knowledge of Internationally Operating Business Professionals*, in "International Journal of

Business Communication" 47 [4], pp. 380-407.
Katan D. 2004, *Translating cultures: An introduction for translators, interpreters and mediators*, St. Jerome, Manchester.
Katan D. 2006, *It's a question of life or death: Cultural differences in advertising private pensions*, in Vasta N. (ed.), *Forms of promotion: Texts, contexts and cultures*, Pàtron Editore, Bologna, pp. 55-80.
Louhiala-Salminen L., Charles M. and Kankaanranta A. 2005, *English as a Lingua Franca in Nordic Corporate Mergers: Two Case Companies*. in "English for Specific Purposes" 24, pp. 401-421.
Loukianenko Wolfe M. 2008, *Different cultures — Different discourses? Rhetorical patterns of business letters by English and Russian speakers*, in Connor U., Nagelhout E. and Rozycki W. (eds.), *Contrastive rhetoric: Reaching to intercultural rhetoric*, John Benjamins, Amsterdam, pp. 87-120.
Manca E. 2016a, *Official Tourist Websites and the Cultural Communication Grammar model: analysing language, visuals, and cultural features*, in "Cultus. The Journal of Intercultural Mediation and Communication" 9 [1], pp. 2-22. http://www.cultusjournal.com/files/Archives/Cultus9_2016_1/Cultus9_2016_Vol%201.pdf (15.10.2018).
Manca E. 2016b, *Persuasion in Tourism Discourse. Methodologies and Models*, Cambridge Scholars Publishing, New Castle upon Tyne.
McSweeney B. 2002, *Hofstede's model of national cultural differences and their consequences: A triumph of faith: A failure of analysis* in "Human Relations" 55, pp. 89-118.
Meeuwesen L., van den Brink-Muinen A. and Hofstede G. 2009. *Can dimensions of national culture predict cross-national differences in medical communication?*, in "Patient Education and Counseling" 75, pp. 58-66.
Nakata C. 2009, *Reflexive Considerations of Culture Theories in Global Marketing*, in Nakata C. (ed.), *Beyond Hofstede. Culture Frameworks for Global Marketing and Management*, Palgrave Macmillan, Basingstoke and New York, pp. 247-277.
Partington A. 2004, *Corpora and discourse, a most congruous beast*, in Partington A., Morley J. and Haarman L. (eds.), *Corpora and Discourse*, Peter Lang, Bern, pp. 11-20.
Piller I. 2009, *Intercultural Communication*, in Bargiela-Chiappini F. (ed.), *The Handbook of Business Discourse*, Edinburgh University Press, Edinburgh, pp. 317-329.
Pym A. 2010, *Website Localizations*, Pre-print text written for the *Oxford Companion to Translation Studies*, pp. 1-14. http://usuaris.tinet.cat/apym/on-line/translation/2009_website_localization_feb.pdf (15.10.2018).
Rayson P. 2009, *Wmatrix: a web-based corpus processing environment*, Computing Department, Lancaster University. http://ucrel.lancs.ac.uk/wmatrix/.
Scott M. 2012, *WordSmith Tools version 6*, Lexical Analysis Software, Stroud.
Seidlhofer B. 2009, *Common ground and different realities: World Englishes and English as a lingua franca*, in "World Englishes" 28 [2], pp. 236-245.
Singh N., Zhao H. and Hu X. 2005, *Analyzing the cultural content of web sites. A cross-national comparison of China, India, Japan, and US*, in "International Marketing Review" 22, pp. 129-146.
Singh N. and Pereira A. 2005, *The culturally customized web site. Customizing web sites for the global marketplace*, Elsevier, Atlanta, GA.

Vishwanath A. 2003, *Comparing Online Information Effects: A CrossCultural Comparison of Online Information and Uncertainty Avoidance*, in "Communication Research" 30 [6], pp. 579-598.

Zhao W., Massey B. L., Murphy J. and Fang, L. 2003, *Cultural dimensions of website design and content*, in "Prometheus" 21, pp. 75-84. https://www.researchgate.net/publication/227623341_Cultural_Dimensions_of_Website_Design_and_Content (15.10.2018).

Websites

German companies
Alpenhein http://www.alpenhain.de/
Ammerland http://www.molkerei-ammerland.de/good-reasons-why-were-at-the-top/document_view
Andechser Natur http://www.andechser-natur.de/
Bauer http://www.bauer-milch.de/en-home/
Bechtel http://www.privatmolkerei-bechtel.de/en/
Bergader http://www.bergader.de/index.php?id=1&L=1
Exquisa http://www.exquisa-international.de/index.php
Goldsteig http://www.goldsteig.de/en.html
Kohroner http://www.kohrener-landmolkerei.com/en/
Petrella http://petrella.de/index.php/home-en.html
Ruecker http://www.ruecker.org/en/
Weidegluck http://www.weideglueck.de/index.php/en/

Spanish companies
Aldonza y Don Ismael http://www.donismael.com/eng/default.htm
Coqueya http://www.coqueya.com/web/english/marco.htm
El Gran Cardenal http://www.elgrancardenal.com/?Idioma=EN
El Pastor http://www.elpastor.com/en/inicio.php
Eusebio Manzano http://www.quesoseusebiomanzano.es/QEM_eng/About_us.html
Garcia Baquero http://www.garciabaquero.com/home.aspx
Los Cameros http://www.loscameros.es/en/
Maxorata http://www.maxorata.es/english/home.htm
Monteverdesa http://www.monteverdesa.com/index.asp?idioma=2
Palancares http://www.quesospalancares.com/en/
Pastoret http://pastoret.com/en/company/
San Vincente http://www.lsanvicente.com/eng/

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 249-267
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p249

RESEARCH ARTICLE ABSTRACTS AS A TOOL FOR DISSEMINATING KNOWLEDGE IN ONLINE LEGAL PUBLICATIONS[1]

MICHELE SALA
UNIVERSITÀ DEGLI STUDI DI BERGAMO

Abstract – In the course of the last few decades web-based communication has facilitated the dissemination of scientific knowledge and has contributed to making material which was primarily targeted to the expert community also available to the wider public. For such a mass of information to be effectively comprehended, some changes became necessary in order to facilitate web-searches and locate relevant material when browsing online archives. In the light of the above, the genre of research article (RA) abstracts in online academic publications seems to be particularly interesting, since its distinctive generic function is to both attract the readers' attention and synthetically anticipate the content of the ensuing RA. This paper compares abstracts in print issues and those available in electronic format of the *Washington Law Review* to see whether and to what extent the concern for piquing readers' curiosity to continue reading the associated RA has affected abstract writing. The focus will be placed particularly on knowledge dissemination (KD) resources such as the labels used to refer to the associated RA, the verbs employed to describe the type of scientific activity performed by the RA and, finally, the metadiscursive markers (namely frame and endophoric markers) used to provide interpretive guidance to readers.

Keywords: abstracts; academic discourse; knowledge dissemination; legal discourse; metadiscourse.

1. Introduction

This paper investigates the language of legal research article abstracts (LRAAs) in online specialized journals from a diachronic perspective. More specifically, our analysis focusses on the different ways in which abstracts textualize references to the associated RA by comparing and contrasting LRAAs published before and after the *Washington Law Review*'s digital shift,

[1] This study is part of a national research project on "Knowledge Dissemination across media in English: continuity and change in discourse strategies, ideologies, and epistemologies", financed by the Italian Ministry of University and Research (PRIN 2015TJ8ZAS).

that is to say, the moment the journal went fully online. The study of abstracts in such contexts appears to be particularly relevant for two related reasons, namely, the purpose of abstract as a research genre (Bhatia 1993; Hyland 2004; Lorés Sanz, Bondi 2014), on the one hand, and the circulation of informative material made available by the electronic medium (Duszak 1997; Gotti 2003; Salvi, Bowker 2015), on the other.

Abstracts are in fact those texts which – while self-standing and autonomous with respect to the associated RA (Hyland 2004) – are primarily meant to point to, synthesize, topicalize and indexicalize its content in a concise, orderly and schematic way (Trawinski 1989; Swales, Feak 2009; Hyland, Bondi 2006). This allows them to be easily read and understood independently from the ensuing RA, once they are cohesive and internally coherent (Salager-Meyer 1990; Ventola 1997). However, RAAs also have an intrinsic promotional function, in that they are typically aimed to elicit the readers' interest towards the content of the associated RA and encourage them to read in full its informative content (Yakhontova 2002; Bordet 2014; Bondi, Lorés Sanz 2014). More precisely, the abstract does not only provide the gist of the RA, but it "selectively sets out the stall, highlighting important information and framing the article that it precedes […] in such a way as to encourage further examination and draw the reader into the more detailed exposition" (Hyland 2004, p. 64).

Promotion owes its (potential) effectiveness not only to the rhetorical strategies employed to codify meanings, but also, and blatantly, to the type of circulation, availability and accessibility of promotional material (Corner 2007; Aronczyk, Powers 2010; Maci 2016). To put it simply, people may become interested in given knowledge objects or reality objects simply because they are exposed to texts promoting or advertising them. Conversely, without access to such promotional materials, one may not even be aware of the existence of such objects, thus possibly never developing any interest or curiosity for them.

In the light of the above, the digital medium appears to be the perfect match for the abstract's promotional function, in that this channel seems to offer plenty of possibilities for the wider circulation of information (Tognini-Bonelli, Del Lungo Camiciotti 2005; Hyland 2009; Campagna *et al.* 2012). As a matter of fact, by exploiting extended web-based archives, it facilitates web-searches and the retrieval of very specific material. As a consequence of this, "knowledge can no longer be hemmed in by neatly-packed and restricted communicative products [since] today's globalized and digital environment has increasingly destructured the élitarian fence of knowledge communication" (Bondi 2015, p. 7).

Given the function of abstracts and the assets of web communication, this analysis seeks to see whether, to what extent and in what ways the textualization of abstracts has changed with the spread of digital communication for this genre to better fit and fully exploit the resources made available by the new channel.

The hypothesis guiding this analysis is that online archives and digital technologies have become important tools for knowledge dissemination (Calsamiglia 2003; Salvi, Bowker 2015) not just for external audiences of laypersons, but also, and in particular, within the scientific community, for the circulation of specialized knowledge among experts (Shinn, Whitley 1985; Abbott 1995; Bucchi 1998; Bondi 2015). If this is the case, this new perspective is likely to have influenced discursive practices and, possibly, also aspects of the epistemology at the basis of specialized communication in the legal domain.

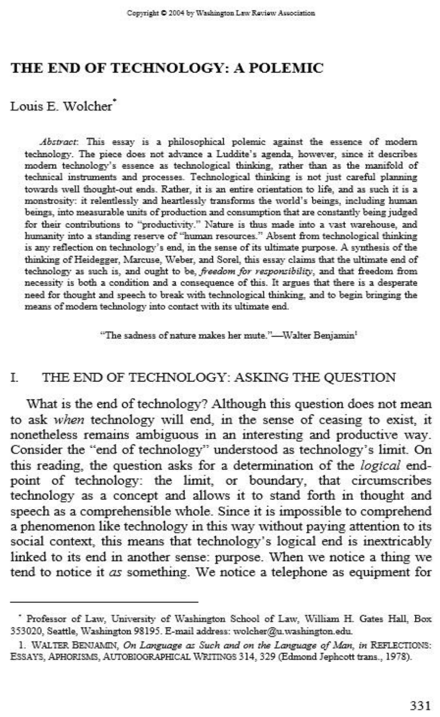
2. Material and methodology

The object of this case study is the *Washington Law Review*, a nationally ranked law review established in 1919, which publishes four issues per year (with an average of 8 RAs per issue) available through open-access databases. More specifically, yearly volumes (containing four issues each) from 1919 to 2007 can be downloaded from a host website[2] as PDF files reproducing the original print version; whereas, from the year 2012 onwards, each single contribution (i.e. abstract and RA) can be accessed separately via a link on the journal homepage.[3] There is a four-year gap in the availability of such material (namely, the issues published in the time span 2008-2011). This gap has been used as a criterion for the distinction of our corpus into two sub-corpora. On the one hand, we have collected the more recent material, covering the six-year span 2012-17 and, by analogy, material from the six year period before the gap, namely the span 2002-2007, has also been selected.

These two sub-corpora appear to be significantly different even at a very superficial level. Material for the 2002-2007 sub-corpus comes from PDF files reproducing the exact text found in the print issues, and each file contains the whole yearly volume (four issues, for a total of about 32 articles per year). For our analysis, each abstract had to be selected and isolated from its context and this sub-corpus amounts to a total of 135 LRAAs (30,037 words, corresponding to an average of 222.5 words per abstract). In these files, abstracts are always placed between the title and the full body of the RA, and in such contexts the abstract can be considered as a part-genre "given the fact that it [is] included within the document and that it could only be read as part of the whole document" (Bordet 2014, p. 132). This is significant from a

[2] At http://digital.law.washington.edu/dspace-law/handle/1773.1/9 (02.04.2018).
[3] At https://www.law.uw.edu/wlr (02.04.2018).

cognitive perspective; in fact, since readers can access the full article together with the abstract, the abstract functions as a sort of concise introduction (Swales, Feak 2009) to the ensuing text rather than a 'teaser' pointing to an associated text which is to be found in a separate section. This factor may explain the fact that in this sub-corpus some abstracts are missing (39 cases), indicating that in this context abstracts, while indeed useful and important, are not necessary



Copyright © 2004 by Washington Law Review Association

THE END OF TECHNOLOGY: A POLEMIC

Louis E. Wolcher*

Abstract: This essay is a philosophical polemic against the essence of modern technology. The piece does not advance a Luddite's agenda, however, since it describes modern technology's essence as technological thinking, rather than as the manifold of technical instruments and processes. Technological thinking is not just careful planning towards well thought-out ends. Rather, it is an entire orientation to life, and as such it is a monstrosity: it relentlessly and heartlessly transforms the world's beings, including human beings, into measurable units of production and consumption that are constantly being judged for their contributions to "productivity." Nature is thus made into a vast warehouse, and humanity into a standing reserve of "human resources." Absent from technological thinking is any reflection on technology's end, in the sense of its ultimate purpose. A synthesis of the thinking of Heidegger, Marcuse, Weber, and Sorel, this essay claims that the ultimate end of technology as such is, and ought to be, *freedom for responsibility*, and that freedom from necessity is both a condition and a consequence of this. It argues that there is a desperate need for thought and speech to break with technological thinking, and to begin bringing the means of modern technology into contact with its ultimate end.

"The sadness of nature makes her mute."—Walter Benjamin[1]

I. THE END OF TECHNOLOGY: ASKING THE QUESTION

What is the end of technology? Although this question does not mean to ask *when* technology will end, in the sense of ceasing to exist, it nonetheless remains ambiguous in an interesting and productive way. Consider the "end of technology" understood as technology's limit. On this reading, the question asks for a determination of the *logical* end-point of technology: the limit, or boundary, that circumscribes technology as a concept and allows it to stand forth in thought and speech as a comprehensible whole. Since it is impossible to comprehend a phenomenon like technology in this way without paying attention to its social context, this means that technology's logical end is inextricably linked to its end in another sense: purpose. When we notice a thing we tend to notice it *as* something. We notice a telephone as equipment for

* Professor of Law, University of Washington School of Law, William H. Gates Hall, Box 353020, Seattle, Washington 98195. E-mail address: wolcher@u.washington.edu.

1. WALTER BENJAMIN, *On Language as Such and on the Language of Man*, *in* REFLECTIONS: ESSAYS, APHORISMS, AUTOBIOGRAPHICAL WRITINGS 314, 329 (Edmond Jephcott trans., 1978).

331

Figure 1
Example of RAA in the 2002-2007 sub-corpus.

The second sub-corpus, instead, contains abstracts that can be accessed directly from the journal homepage via a link, that is, separately from the associated RA, which, in its turn, is retrievable via a different link (namely, the hyperlink consisting of the RA title, see figure below).

Figure 2
Example of RAAs in the 2012-2017 sub-corpus.

A total of 191 abstracts have been collected for the 2012-2017 sub-corpus (totalling 45,648 words, with an average of 239 words per abstract). In this case, as well, 23 abstracts are missing but, due to the constraints of the medium – namely, the link labelled 'View Abstracts' (see Figure 2) – they have been replaced by 'Excerpts', that is, citations taken verbatim from the related RA (see Figure 3, where the substituting text is introduced by the terms 'Excerpt' in parentheses).

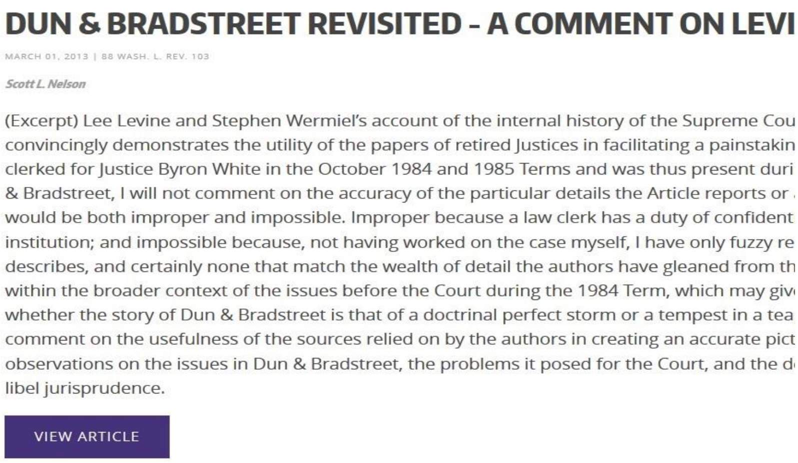


Figure 3
Example of 'excerpt' replacing RAAs in the 2012-2017 sub-corpus.

By comparing the textualizations found in the two sub-corpora (Figure 1 and Figure 2) a noticeable difference can be seen between them: the print version (or, more properly, the digital version replicating the print version, in the 2002-

2007 sub-corpus) allows for more structural freedom (namely, for the possible absence of abstracts), whereas the digital format in the 2012-17 sub-corpus, due to its organizational structure, is much more fixed, requiring some 'texts' to correspond to the link 'View Abstract', be it an actual abstract or a text replacing or filling in for it.

As to the quantitative and qualitative analysis of these materials, the focus has been placed on three different aspects which may indicate possible changes or trends of variation in the way abstracts conceptualize, anticipate and picture the associated RA and the type of discussion to be found there. Such differences, in fact, are likely to affect expectations on the part of the readers and the way they will approach and process the content of the associated text. The parameters considered here are:

- the labels which are employed to refer to the associated RA;
- the verbs used to describe the scientific activity carried out by the RA, that is, how the information is going to be presented;
- the use of frame and endophoric markers meant to anticipate the structural and textual organization of the associated text, thus functioning as cognitive facilitators or interpretive frameworks.

3. Results

3.1. Labels

The terms used to introduce the associated text are very relevant in that the act of classifying, identifying or associating a given RA with respect to recognizable research genres necessarily anticipates some information, not just about the content, but possibly also as to how the content is going to be dealt with, both in textual, cognitive and argumentative terms, hence presupposing the type of competence that might be required to fully comprehend the ensuing text, and the interpretive stance and approach to be adopted for the processing of the meaning. In total, six different terms have been found in LRAAs to refer to the associated RAs, namely: *article, comment, essay, note, paper* and *response.* Even though some of these terms are quasi-synonymic, a major distinction between two macro-groups can easily be introduced, which is relevant in terms of presupposition about the type of content and the way is it going to be dealt with in the RA. On the one hand, we found the labels *article, essay* and *paper,* on the other, c*omment, note,* and *response.*

Terms like *article, essay* and *paper* refer to genres which are self-standing and self-sufficient in terms of content and informative structure, in that everything that is needed to understand the discussion (data, premises, procedures, principles and parameters for interpretation) is expected to be provided and clearly worded out. Therefore, these labels seem to point to or presuppose texts which are possibly – and at the same time – textually articulated and interpretively demanding (i.e. relatively long, informatively dense, etc.), on the one hand, but also cognitively manageable and non-problematic to follow and understand, in that interpretation and evaluation is expect to be prepared by the exposition and explanation of the data to be interpreted and of the principles by which to evaluate them. Some examples of such uses can be seen in the extracts below:

1) [This] paper briefly addresses how management agencies might be encouraged to adopt such an approach. (07/17)

2) [This] Essay concludes that a "democratic competence" approach might provide a more coherent theoretical underpinning for according constitutional protection to newsgathering. (12/15)

The labels *comment*, *note* and *response*, instead, presuppose some form of thematic interdependence between the text at stake and (pieces of) existing knowledge. Hence, texts being referred to through these labels are represented as not being fully self-standing, but as hinging on some form of interdiscursivity, 'dialogicality' or intertextuality (Salvi 2015), since they imply the interaction of different 'voices' (namely, the voice of the writer and the one of the scholar being commented upon or responded to, cf. Bakhtin 1981). In other terms, these labels embody a responsive act rather than an act initiating an informative offer. For this reason, they seem to point to an ensuing text which is expected to be little articulated or cognitively demanding but, at the same time, which is likely to require some threshold competence about the knowledge object being commented upon. They appear to presuppose expert readers, and, more specifically, readers who have some understanding about the informative gap which is going to be filled by the associated RA. Therefore, these labels emphasize collegiality, group-membership and, for the very same reason, they may be face-threatening for non-expert readers, thus possibly functioning as gate-keeping resources, as can be observed in the following examples

3) The Comment then reviews the psychological and social science research. (17/09)

4) This brief response to the work of Professors Omri Ben-Shahr and Carl Schneider on mandated disclosure regimes investigates the normative criteria underlying their claim that those regimes are failures. (13/14)

The distribution of the two groups of labels is organized in the following Table, and is expressed both in absolute terms and in terms of percentage. More specifically, these results do not express the total occurrence of a given label in each RA, since this quantitative piece of evidence would not be relevant for this analysis, but rather count the RAs in which a given label has been found.

	2002-2007	%		2012-2017	%
article	35/135	27		89/191	47
essay	1/135	1		3/191	2
paper	1/135	1		2/191	1
subtotal		***29***			***50***
comment	58/135	42		60/191	31
note	14/135	10		3/191	2
response	-	-		1/191	0.5
subtotal		***52***			***34***

Table 1
Distribution of labels to refer to RAs in the two sub-corpora.

As we can see, abstracts in the 2002-2007 sub-corpus clearly privilege label indicating interdiscursivity, and implying collegiality and group membership, by labelling the associated RAs as *comments*, *notes* or *responses* (52%), whereas the 2012-2017 sub-corpus reverses this tendency by resorting to labels emphasizing the semantic autonomy of the associated texts. Even at this early stage of the investigation, a major difference appears to characterize the two sub-corpora, both in terms of expectations and approach towards the content of the RA: while print abstracts rhetorically presuppose readers with discipline-related expertise for them to comprehend the ensuing text, digital abstracts discursively imply the possibility for readers to find in the associated text anything that is needed to process it, that is, the discursive framework by which to approach its content, and the main interpretive guidelines by which to navigate through its textualization.

3.2. The verbal expressions

The second step in our analysis consists in examining the verbs used in collocates with the labels discussed above and meant to anticipate the type of act or scientific activity performed by the associated RAs. Verbs in such clusters can be grouped into three macro-categories with respect to their lexical meaning, and the presuppositions that such markers activate.

The first of such groupings is represented by those verbs pointing to the type of analysis carried out in the associated text. This set is mainly made up of *research* verbs and those verbal forms indicating how the presentation is organized, how the material is dealt with, how the investigation is carried out and how it is sequenced, thus evidencing procedural and systematic aspects of the discussion, through verbs like *address, analyse, discuss, explain, focus*, etc., as in the following cases:

5) This Article analyzes the benefits of using "stories" to teach law. (13/24)

6) This Comment presents newly produced data sets. (17/09)

The second group consists of verbs indicating the *purpose* of the discussion found in the associated text, emphasizing its argumentative nature and presupposing that its content is not just the objective and unbiased representation of informative material but rather its interpretation according to specific parameters. These readings are codified by verbs such as *argue*, *contend, conclude*, *challenge*, *demonstrate*, *evaluate*, etc. By the use of these markers, the RA is presented as being explicitly meant to persuade the readers rather than inform them, to convince them rather than just provide evidence for readers to measure, evaluate, and interpret. In such cases, of course, informativeness is not missing or marginalized, but is rather the epiphenomenon of argumentation. Examples of these verbal formulations can be found in the following extracts:

7) This Article contends that section 2 influences the scope of congressional authority. (14/15)

8) This Comment argues that courts should award damages to tribes. (04/35)

The third group includes verbs which, although pertaining to the lexical category of research verbs, do not just refer to representational, organizational or textual aspects of the discussion at stake, but specifically point to the *type of contribution* the RA is intended to bring to existing knowledge, thus rhetorically anticipating to what extent and through which resources the associate RA is going to expand, extend, or update the community knowledge. Instances of these markers are verbs like *contribute*, *develop, fill [gaps]*, *offer*, *reveal,* etc., that can also be seen in the extracts below:

9) [This] Article provides a comprehensive survey of state whistleblowing laws and suggests changes to federal and state law to fill the gaps that remain after Sarbanes–Oxley. (04/32)

10) This Article is the first to comprehensively consider the intersection of procurement and local surveillance policy making. (16/31)

The main functions of the three groups outlined above can be synthesized as follows: verbs in the first group are meant to express an act of *speculation*; those in the second, an act designed to both *prove a specific point* and *persuade* about its validity; those in the third group are meant to indicate and stress the *originality, novelty* and *usefulness of the contribution* with respect to existing knowledge. Frequencies and distributions of these resources are organized in the following Tables, expressed in normalized figures (calculated per 10,000 words).

	2002-2007		2012-2017	
RESEARCH	article / essay / paper	comment / note / response	article / essay / paper	comment / note / response
address	0.6	0.3	1.3	1.1
analyse	0.6	0.6	1.1	1.1
assess			0.9	
consider	0.3		1.3	
describe	0.3		0.4	0.2
discuss	0.6		1.1	0.4
draw			0.9	
examine		0.6	3.5	0.9
explain	0.3		0.6	0.2
explore	0.6	0.3	2.1	1.7
focus	0.3		0.4	0.4
identify	0.3		1.5	0.2
present			0.9	0.6
propose	2.0	3.3	2.4	0.4
review	0.3		0.6	0.6
show	0.3		1.1	
suggest	0.6	0.3	0.2	0.6
subtotal	**10.0**	**5.6**	**21.6**	**8.8**
TOTAL	**15.6**		**30.5**	

Table 2
Frequencies and distribution of *research* verbs.

	2002-2007		2012-2017	
ARGUMENTATION	article / essay / paper	comment / note / response	article / essay / paper	comment / note / response
argue	2.6	21.3	5.2	8.9
challenge			0.9	
conclude	2.0	0.6	2.8	0.4
contend	0.3		0.4	0.2
demonstrate			0.9	
expose	0.3		0.4	0.2
evaluate	0.6		0.2	0.2
recommend	0.9			0.4
subtotal	7.0	21.9	11.0	10.5
TOTAL	**28.9**		**30.4**	

Table 3
Frequencies and distribution of *argumentation* verbs.

	2002-2007		2012-2017	
CONTRIBUTION	article / essay / paper	comment / note / response	article / essay / paper	comment / note / response
contribute	0.3		0.4	
develop	0.6		0.6	
fill [gaps]			0.6	
offer	0.9		1.1	0.2
provide	0.6		1.3	0.2
reveal	0.3		1.3	
* is the first to			2.8	
subtotal	**2.9**		**8.3**	**0.4**
TOTAL	**2.9**		**8.7**	

Table 4
Frequencies and distribution of ‘*contribution’* verbs.

In order to facilitate synoptical comparison between the three categories, the frequencies of these verbs can be observed in Table 5.

	2002-2007	2012-2017	TOTAL
research m.	15.6	30.5	**46.1**
argumentation m.	28.9	30.4	**59.3**
contribution m.	2.9	8.7	**11.6**
TOTAL	**47.4**	**69.6**	

Table 5
Frequencies and distribution of *research*, *argumentation* and ‘*contribution’* verbs.

By observing the total frequencies, that is, the last line at the bottom of Table 5, it is possible to see a noticeable increase in the use of these resources over time, as if in digital communication displaying these verbs with the function of metatextual markers meant to anticipate and explicitate the type of activity performed by the RA were an effective way of promoting it and eliciting the reader’s interest.

If we then observe the second line and compare occurrences in the two sub-corpora, we notice that argumentation markers are a constant feature over time. Argumentation appears to be the preferred and conventional way of presenting domain-specific meanings, and this seems to be in line with the epistemology at the heart of the discipline, in that legal studies focus on the interpretation of practical cases in the light of abstract principles, and argumentation is an effective way of codifying interpretation (Toulmin 1958; van Eemeren, Grootendorst 2004). However, what varies considerably between print and digital abstracts is the occurrence of research and contribution markers (which are respectively two and three times more frequent in the 2012-2017 sub-corpus). By resorting to such resources, digital abstracts indeed attribute another dimension to the associated text: they represent the RA as a fully-fledged research genre carrying out a recognizable form of scientific activity. If only from a purely discursive point of view, this contributes to extending the pragmatic function of RA as a communicative tool for knowledge dissemination, and also extending the range of discursive practices available to the disciplinary community (i.e. introducing research-based acts besides traditional argumentation-based ones) for the transmission of domain-specific contents.

3.3. Frame and endophoric markers

The last stage of this analysis focusses on frame and endophoric markers, that is, those resources which are used to "signal text boundaries or […] to sequence, label, predict and shift arguments, making the discourse clear to readers" (Hyland 2005, p. 51). In other words, these resources are exploited as cognitive facilitators to signal how the associated RA is going to be structured and to mark textual cohesion and coherence, thus helping the reader 'visualize' the stages, the steps and the sections to be found in the associated texts, and how they are combined in order to make a point. For this purpose, we have analysed collocates with the terms discussed in section 3.1. above, namely, the labels *article*, *essay*, *paper*, *comment*, *note*, and *response*. After examining all occurrences, we have observed that the terms which are associated to such labels can be distinguished according to their metadiscursive function into two categories, namely, markers of *formal articulation* and markers of *content articulation*. Formal articulation is expressed by either *ordering markers* (i.e. *first, second, third*, etc.) or *sequencing markers*, anticipating textual structure and sequencing between the various parts of the RA (i.e. *next*, *then*, *finally*, etc.), as the examples below show:

11) This Comment first identifies and explains the different meanings attached to loss of chance. (14/21)

12) [This] Article then scrutinizes the background legal doctrine framing this debate. (05/22)

Content articulation markers are formulations expressing addition (like *also* and *further*) or contrast (such as *instead*), as can be seen in the following extracts:

13) Furthermore, the proposals in this Article provide a blueprint for advocates (15/27)

14) This Article instead suggests an analysis of ADA hostile environment actions (02/11)

The frequency and distribution of these markers is indicated in normalized terms in the table below:

	2002-2007	2012-2017
Formal art. markers: ordering		
first / firstly	0.9	4.1
second / secondly	2.0	3.2
third / thirdly	0.6	1.1
subtotal	***3.6***	***8.5***
Formal art. markers: sequencing		
finally	1.3	4.3
next	0.6	0.6
then	1.3	4.6
ultimately		0.4
subtotal	***3.3***	***10.0***
Content art. markers		
also	1.3	3.0
further	2.0	1.1
instead	0.3	0.9
subtotal	***3.6***	***5.0***
TOTAL	**11.0**	**23.6**

Table 6
Distribution of text articulation markers.

If we consider the frequencies in the 2002-2007 sub-corpus, we notice that the use of the three types of resource is quite balanced (3.6 occurrences for ordering markers, 3.3 for sequencing markers and 3.6 for content articulation markers). In print LRAAs, these indications seem to share the same level of rhetorical relevance and there does not seem to be any preferred or more conventional way to mark text articulation. Even a cursory glance at frequencies in the 2012-2017 sub-corpus, instead, reveals that these strategies are noticeably more used in digital abstracts (8.5 occurrences for ordering markers, 10.0 for sequencing markers and 5.0 for content articulation markers). Metadiscursively anticipating elements of text articulation – i.e. combining ideational contents and structural or organizational indications – seems thus to be an effective way of representing the associated RA in electronic abstract writing, whereby to attract readers' interest by also providing, if in a very schematic way, an interpretive framework for the processing of the text. A more detailed analysis of the occurrences in the digital sub-corpus indicates a marked preference for formal articulation markers (of both ordering and sequencing) over content articulation markers, the frequency of the former pair almost doubling that of the latter. This suggests that in digital LRAAs representing the scaffolding, the main tenets and cognitive hinges of the ensuing discussion (*first, next, then,* etc.) may attract readers more effectively than simply indicating general content-based and additive relations between pieces of information (i.e. *also, instead,* etc.). In other words, structural indications are (presented as being) more appealing to digital audiences than stacks or chunks of informative material.

4. Concluding remarks

The W*ashington Law Review* in its electronic format, if accessible also to laypersons, remains a specialized text primarily targeted to experts, as is the case of plenty of other specialized publications that have undergone a similar process of digital democratization, since this operation does not necessarily coincide with a top-down process of simplification (Bondi 2015, p. 8). As a matter of fact, the electronic version of the journal would not qualify as a form of popularization in that, if cognitively transparent and potentially manageable, it does not contain attractors or facilitators to capture the layperson's curiosity – like visuals, simplified language, explanations or exemplification, as is instead the case, for instance, of popularized science journals (cf. Calsamiglia 2003; Gotti 2013). In this respect, it is possible to claim that the digital resources are here employed as tools of knowledge dissemination primarily for the expert community – in order to spread as extensively, comprehensively and readily as possible pieces of specialized knowledge among legal scholars and practitioners, so as "to reach a vast number of colleagues rapidly by sending them 'coded messages' without having to conform to the times and constraints of specialist communication" (Bucchi 1998, p. 12).

In this sense, the changes that have been observed in the previous sections are particularly interesting. Since they cannot merely be viewed as an attempt at updating the language of LRAAs to current standards of web communication, these trends seem to indicate a gradual shift in the epistemology of web-based communication of legal matters, at least concerning how legal RAs and their way of presenting contents are to be considered.

As a matter of fact, abstracts in the print version of the W*ashington Law Review* picture RAs in ways by which it is possible to recognize them as typical, highly-conventionalized, authoritative and reliable argumentative texts about legal topics, reflecting the main functions and requirements of legal RAs as a genre. This genre has a clear argumentative character, a gate-keeping quality and an emphasized interpersonal dimension which is meant to foster persuasiveness by means of modalization, evaluation and interactional marking (Goodrich 1987; Fish 1989; Gotti, Williams 2010; Breeze *et al.* 2014).

Legal studies are argumentative in nature since they interpret contextual situations with respect to general paradigms or precedents, and argumentation (rather than exposition of objective data, measuring or experimental evidence, hypothesis testing, etc.) is the most appropriate mode for this type of activity. This attitude is clearly reflected in the print abstracts collected in the 2002-2007 sub-corpus, where argumentation markers can be found both in the verbs which are employed to point to the type of activity performed by the associated RAs – that is represented as the outcome of the act of arguing, contending, challenging, etc. – and also in the type of text-articulation connectors exploited to signal content organization – stressing meaning relations (i.e. *instead, also*), rather than signalling stages (*first, then, next,* etc.), thus positing an argumentative, content-based framework in order to favour specific interpretations.

As regards gate-keeping, the *Washington Law Review* discusses issues which are mainly relevant or applicable to Anglo-American contexts, therefore based on the Common Law system, which is based on the doctrine of precedent; hence some forms of interdiscursivity are implicit in the disciplinary discourse. More precisely, this discursive interconnectedness is part of the epistemology of the domain and is reflected in its conventional discursive practices.

Print LRAAs in the 2002-2007 sub-corpus – aligning, as we have seen, with traditional and conventional discursive requirements and constraints – expressly codify and mark such intertextuality and interdiscursivity by labelling the associated RAs as *comment*, *note* or *response*, thus highlighting their being turns in a dialogistic continuum, represented by the (dominant) views of the community, rather than as self-contained and independent informative offers. By representing RAs in these terms, print abstracts presuppose both group membership and gate-keeping, since discursive interconnectedness hinges on in-group conventions which can only be appreciated by expert users.

The particularities and peculiarities that have been observed in the textualization of digital LRAAs in the 2012-2017 sub-corpus, instead, may be indicative of a gradual change in the epistemology of legal studies. In fact, while maintaining the prosody which typifies print abstracts, digital abstracts tend to introduce the associated RA in a significantly different way, noticeably by minimizing and concealing gate-keeping. Prosodic continuity with traditional abstract writing is testified by the fact that the occurrences of rhetorical and metadiscursive resources in the 2012-2017 sub-corpus (namely verbs and endophoric/frame markers) never decrease if compared to the 2002-2007 sub-corpus. The only element of variation is to be found in the varying degree of their increment (as can be seen especially in Tables 5 and 6). In this respect, an interesting piece of evidence is the fact that, among such markers, those whose increase is comparatively more contained are precisely those which are instead more frequent and emphasized in print LRAAs, namely those pointing to interdiscursivity and argumentation. As a consequence, while print abstracts represent RAs as produced by expert arguers negotiating their point of view with a competent audience, digital LRAAs introduced RAs as the product of expert researchers contributing to extending disciplinary knowledge with a solid and articulated piece of research which, if primarily targeted to a specialized audience, can be potentially accessible also to lay readers having some interest in the topics being discussed.

As a matter of fact, digital abstracts tend to codify references to the associated RA as a semantically autonomous text, lexicalizing it through labels (*article*, *essay*, *paper*) which point to its internal coherence as if to limit as much as possible any gate-keeping potential. Secondly, the RAs keep being represented as argumentative texts but in the 2012-2107 sub-corpus there is a more marked metatextual emphasis on research-related acts (through verbs like *study, analysis, explore, investigate,* etc.) and on the novelty and originality of the contribution (through verbs such as *contribute, develop, offer, is the first to,* etc.). The promotion of legal RAs, and consequently the dissemination of their contents, appears to benefit from linguistically introducing them as solid pieces of scientific research, whose validity can be appreciated even outside the community of reference, in that they produce results which are (anticipated as being) relevant, innovative and original and, therefore, which are bound to contribute to the advancement of knowledge.

From another angle, the trends observed in this study indicate that in the 2012-2017 sub-corpus the promotional function of LRAAs towards RAs is also performed at the level of cognitive accessibility or manageability. This is reflected in a progressive movement towards limiting and circumscribing traditional or conventionalized traits (i.e. emphasis on argumentation and interdiscursivity). Discursive conventions are indeed strategic short-cuts toward understanding, but they are only effective to experts. More precisely, they require a threshold level of disciplinary competence or, notably, they presuppose RAs being assessed through a very specialized perspective and stance, and this way of approaching the interpretation of a text may be cognitively demanding (Bucchi 1998; Hyland 2004). On this basis, the discursive democratization observed in this context does not seem to be primarily necessitated by the need to attract lay audiences, but rather to make specialized contexts appealing and easy to process for legal experts. As a matter of fact, given the growing amount of specialized material retrievable from electronic sources (Jaime Sisó 2009) and the mass of informative material to choose from, digital abstracts are aimed at facilitating the operation of scanning-reading, of locating keywords (labels and verbs) revealing the type of scientific activity to be found in the text, as well as the main tenets and various stages through which the discussion of a given content is organized (*first, second, then,* etc.).

As a consequence, digital LRAA writing – hence, also the representation of legal contents and of the genres designed for their discussion – seems to have adjusted to the possibilities of the electronic medium, on the one hand, and, on the other, also to new audiences of expert readers, who resort to web-searches with specific expectations and needs, which are different from those activated when processing traditional genres and conventionalized disciplinary discourses.

Bionote: Michele Sala, PhD (University of Bergamo), MA (Youngstown State University, Ohio), is an associate professor in English Language and Translation at the University of Bergamo. He is a member of AIA (Associazione Italiana di Anglistica), CERLIS (Centro di Ricerca sui Linguaggi Specialistici), CLAVIER (Corpus and Language Variation in English Research Group) and a member of the scientific board of the CERLIS Series (international peer-reviewed volumes on specialized languages). His research activity deals with language for specific purposes and, more specifically, the application of genre and discourse analytical methods to the corpus-based study of specialized texts in the domain of legal research, medicine and applied linguistics (*English Language across Contexts, Media and Modes* [co-authored with S. Consonni], *Genre Variation in Academic Communication. Emerging Disciplinary Trends*, [co-edited with S.M. Maci]), *The Language of Medicine: Science, Practice and Academia* and *Insights into Medical Communication* [both co-edited with M. Gotti and S.M. Maci].

Author's address: michele.sala@unibg.it

References

Abbott A. 1995, *Things of Boundaries*, in “Social Research” 62 [4], pp. 857-82.
Aronczyk M. and Power D. (eds.) 2010, *Blowing Up the Brand: Critical Perspectives on Promotional Culture*, Peter Lang, Bern.
Bakhtin M. 1981, *The Dialogic Imagination*, University of Texas Press, Austin.
Bhatia V. 1993, *Analyzing Genre: Language Use in Professional Settings*, Longman, London.
Bondi M. 2015, *Preface*, in Salvi R. and Bowker J. (eds.), 2015, *The Dissemination of Contemporary Knowledge in English. Genres, Discourse Strategies and Professional Practices*, Peter Lang, Bern, pp. 7-9.
Bondi M and Lorés Sanz R. (eds.) 2014, *Abstracts in Academic Discourse. Variation and Change*, Peter Lang, Bern,
Bordet G. 2014, *Influence of Collocational Variations on Making the PhD Abstract an Effective ‘Would-be Insider’ Self-Promotional Tool*, in Bondi M and Lorés Sanz R. (eds.), *Abstracts in Academic Discourse. Variation and Change*, Peter Lang, Bern, pp. 131-160.
Breeze R., Gotti M. and Sancho Guinda C. (eds.) 2014, *Interpersonality in Legal Genres*, Peter Lang, Bern.
Bucchi M. 1998, *Science and the Media: Alternative Routes in Scientific Communication*, Routledge, Oxon/New York.
Calsamiglia H. 2003, *Popularization Discourse*, in “Discourse Studies” 5 [2], pp. 139-146.
Campagna S., Garzone G., Ilie C. and Rowley-Jolivet E. (eds.) 2012, *Evolving Genres in Web-mediated Communication*, Peter Lang, Bern.
Corner J. 2007, *Mediated Politics, Promotional Culture and the Idea of `Propaganda'*, in “Media, Culture and Society” 29 [4], pp. 669-677.
Duszak A. 1997, *Culture and Styles of Academic Discourse*, Mouton De Gruyter, Berlin.
Fish S. 1989, *Doing What Comes Naturally: Change, Rhetoric, and the Practice of Theory in Literary and Legal Studies*, Duke University Press, Durham/London.
Goodrich P. 1987, *Legal Discourse. Studies in Linguistics, Rhetoric and Legal Analysis*, Macmillan, London.
Gotti M. 2003, *Specialized Discourse*, Peter Lang, Bern.
Gotti M. 2013, *The Analysis of Popularization Discourse: Conceptual Changes and Methodological Evolutions*, in Kermas S. and Christiansen T. (eds.), *The Popularization of Specialized Discourse and Knowledge across Communities and Cultures*, Edipuglia, Bari, pp. 9-13.
Gotti M, and Williams C. (eds.) 2010, *Legal Discourse across Languages and Cultures*, Peter Lang, Bern.
Hyland K. 2004, *Disciplinary Discourses. Social Interactions in Academic Writing*, Longman, London.
Hyland K. 2005, *Metadiscourse*, Continuum, London.
Hyland K. 2009, *Academic Discourse: English in a Global Context*, Continuum, London.
Hyland K. and Bondi M. (eds.) 2006, *Academic Discourse across Disciplines*, Peter Lang, Bern.
Jaime Sisó M. 2009, *Anticipating Conclusions in Biomedical Research Article Titles as a Persuasive Journalistic Strategy to Attract Busy Readers*, in “Miscelánea: A Journal of English and American Studies” 39, pp. 29-51.
Lorés Sanz R. and Bondi M. 2014, *Introduction*, in Bondi M and Lorés Sanz R. (eds.) *Abstracts in Academic Discourse. Variation and Change*, Peter Lang, Bern, pp. 9-17.
Maci S. 2016, *Tourism Discourse*, ECIG, Genova.

Salager-Meyer F. 1990, *Discursive Flaws in Medical English Abstracts: A Genre Analysis per Research and Text Type*, in “Text” 10 [4], pp. 365-384.

Salvi R. 2015, *Re-contextualizing Specialized English: From Legislation to Business*, in Salvi R. and Bowker J. (eds.), *The Dissemination of Contemporary Knowledge in English. Genres, Discourse Strategies and Professional Practices*, Peter Lang, Bern, pp.19-45.

Salvi R. and Bowker J. (eds.) 2015, *The Dissemination of Contemporary Knowledge in English. Genres, Discourse Strategies and Professional Practices*, Peter Lang, Bern.

Shinn T. and Whitley R. 1985, *Expository Science: Forms and Functions of Popularisation*, D. Reidel Publishing Company, Dordrecht/Boston/Lancaster.

Swales J. and Feak C. 2009, *Abstracts and the Writing of Abstracts*, The University of Michigan Press, Ann Arbor.

Tognini-Bonelli E. and Del Lungo Camiciotti G. 2005, *Strategies in Academic Discourse*, John Benjamins, Amsterdam.

Toulmin S. 1958, *The Uses of Argument*, Cambridge University Press, Cambridge.

Trawinski B. 1989, *A Methodology for Writing Problem Structured Abstracts*, in “Information Processing and Management” 25 [6], pp. 693-702.

van Eemeren F and Grootendorst R. 2004, *A Systematic Theory of Argumentation. The Pragma-dialectical Approach*, Cambridge University Press, Cambridge.

Ventola E. 1997, *Abstracts as an Object of Linguistic Study*, in Cmejrkova S., Danes F. and Havolova E. (eds.), *Writing vs Speaking*, Gunter Narr, Tubingen, pp. 333-352.

Yakhontova T. 2002, *‘Selling’ or ‘Telling’? The Issue of Cultural Variation in Research Genres*, in Flowerdew J. (ed.), *Academic Discourse*, Pearson Education, Harlow, pp. 216-232.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 269-289
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p269

THE DISCOURSE OF BIOETHICS IN THE ECtHR CASE-LAW
A corpus-based perspective

JEKATERINA NIKITINA
UNIVERSITY OF MILAN

Abstract – This corpus-based study investigates linguistic practices and strategies of dealing with issues of bioethics in the case-law of the European Court of Human Rights ("ECtHR"). The study aims at researching the linguistic strategies of knowledge production and entextualisation in 92 judgments of the ECtHR, focusing on the interaction between legal and bioethical term-related phraseological units – i.e. multi-word terms and term-embedding collocations with a verb – their structure and distributional patterns. Recurrent phraseological units are identified and analysed using methods of corpus linguistics and the theoretical framework of specialised phraseology. The study pursues general descriptive goals and aims at researching the balance and intersection between bioethically charged phraseology and legal phraseology. The main focus is placed on the analysis of typicality of patterning, expressed in terms of domain specificity, association score and log-likelihood.

Keywords: bioethics; ECtHR judgments; specialised phraseology; multi-word terms; term-embedding collocations.

1. Bioethics and the ECtHR

The concept of bioethics eludes clear-cut definitions as this highly specialised field, standing at the crossroads between biomedical, ethical, legal and technological areas, is constantly changing along with the developments in the modern world. Most definitions of bioethics stress its interdisciplinary and changing nature. For instance, Article 1 of the UN Declaration on Bioethics and Human Rights (2005) defines bioethics as "ethical issues related to medicine, life sciences and associated technologies as applied to human beings, taking into account their social, legal and environmental dimensions". Similarly, the Research report on Bioethics of the Council of Europe defines it as "protection of the human being (his/her human rights and in particular human dignity) in the context of the development of biomedical sciences" (2016, p. 4).

Although there are multiple legal instruments dealing with specific issues related to the bioethical field – such as surrogacy laws, assisted suicide acts, etc. – there are only two all-encompassing international legal documents dedicated to the field in general. These are *Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine* ("the Oviedo Convention", 1997) and the *Universal Declaration on Bioethics and Human Rights* (2005).

The European Court of Human Rights ("ECtHR") is the only court which may interpret the Oviedo Convention (Oviedo Convention, art. 29) and which guarantees judicial protection of bioethically relevant rights of a single individual on the supranational level. Its case-law exerts a significant influence on the interpretation and knowledge dissemination of many controversial concepts of bioethics, which are subdivided in several categories in the Court's case-law: physician assisted suicide, consent to medical examination or treatment, ethical issues concerning HIV, retention of biodata (including DNA and fingerprint samples), medically assisted procreation, reproductive rights (including prenatal diagnosis and the right to a legal abortion), liability of health professionals, right to know one's biological identity, transgender issues and rights of children born out of surrogacy agreements (Research report on Bioethics of the Council of Europe 2016, p. 4).

The ECtHR rules on violations of human rights coming from one or several of the 47 Member States of the Council of Europe ("CoE"), with 37 different languages, making it a fertile ground for cross-linguistic and cross-systemic contacts. At the same time, the Court – represented by 47 judges elected in respect of the CoE Member States sitting in different judicial formations – operates with only two official languages, English and French (Rule 34, Rules of Court). Consequently, from the linguistic point of view the case-law of the ECtHR cannot but present features of translation and/or second language production (Nikitina 2018).

In addition to linguistic, translational and legal challenges, the ECtHR judgments dealing with bioethical issues represent a locus of interdisciplinary contact, or to quote Myers "a terrain of competing discourses and practices" (2003, p. 267). The ECtHR judges have to overcome the difficulty of operating with different kinds of knowledge, inserting scientifically charged notions in a rather crystallised structure of a judgment embedded in a legal context. Consequently, these judgments become a hybrid form of knowledge production. Calsamiglia and van Dijk (2004, p. 373) claim that "each kind of knowledge […] has its own format". It could be hypothesised that bioethical knowledge, in light of its interdisciplinary nature, is prone to demonstrate mixed formats incorporated from a range of perspectives and fields.

This paper builds upon the theoretical notion of *entextualisation*, originally elaborated by anthropologists and used by linguists to address the issue of knowledge dissemination and popularisation. Garzone (2014, p. 79) defines entextualisation as "the extraction of meaning from one discourse and consequent insertion of that meaning into another discourse through a process of de-contextualization or 'decentering' and its 're-centering' in another context". This paper looks at how legal and bioethical discourses interact at the lexico-grammatical level, attempting to trace a relationship between the language component (language as communication) and language as vehicle for conveying disciplinary knowledge (Bhatia 2014 [2004], p. 204) through specialised phraseological patterns that revolve around terms and act as "depositories of knowledge" (Sager 1998, p. 259). The basic presumption is that bioethical patterns could present certain traits of legal contamination on account of the already existing legal dimension of bioethics and the immediate legal context of the ECtHR judgment, where these are placed.

2. Specialised phraseology and corpus linguistics

The topic of languages for special or specific purposes (LSP) is widely discussed in the relevant literature, starting from a lexicalised perception of LSPs in 1970s-1980s (e.g. Widdowson 1979, p. 24) to a more recent reanalysis of LSPs that covers also such aspects as specialised phraseology. While LSP phraseology has gained popularity in academic circles in recent years, it is still placed at the periphery of phraseological studies (Kjær 2007, p. 506). In this paper specialised phraseology is understood as prefabricated lexico-grammatical patterns that contribute to the make-up of bioethically-charged ECtHR judgments "at the macrostructural and microstructural level, providing a stable matrix to be filled with details" (Biel 2014a, p. 36).

In general, phraseology interweaves aspects of four disciplines: semantics, morphology, syntax and discourse (Granger, Paquot 2008, p. 30) and represents a dynamic field in constant development. The combination of the phraseological paradigm with the methodology of corpus linguistics is particularly appropriate for the analysis of legal discourse on bioethics – another dynamic field – because corpus linguistics enlarges the research focus "beyond the single word as the basic semantic unit" (Teubert 2002, p. 212). Phraseological analysis of specialised corpora contributes to understanding "the kinds of language data which particular communities of users might encounter and which will inform their use" (Hyland 2008, p. 8). In line with recent research orientations, this study aims at researching the specialised discourse of bioethics in the ECtHR case-law through the paradigm of specialised phraseology.

It has been acknowledged that phraseological units in legal texts tend to cluster around terms, forming a continuum with fuzzy borders (Scarpa *et al*. 2014, p. 75), and terms are usually those elements that could denote a specificity to a certain specialised domain. The interaction between terms and phrases thus is the key to understanding a discipline. Hence, I decided to focus this research on the so-called *term-related phraseological units* (Nikitina 2018, p. 315), i.e. those phraseological units that feature a term. The primary focus of this paper is placed on a special category of multi-word units, which could be classified both as terminological and phraseological. These multi-word units are known in the literature under a variety of labels, such as *terminological phrase* (Kjær 1990), *terminological phraseme* (Meyer, Mackintosh 1994), *multi-word terminological phrase* (Bergenholtz, Tarp 1995), *multi-word term* (Kjær 2007, p. 509) or *term-forming pattern* (Biel 2014b, p. 180), i.e. those terms that consist of more than one word. It is presumed that multi-word terms have the potential to carry highly domain-specific meanings, as the combination of various single words increases the specificity of knowledge charge. In this regard, I draw on Sinclair's interpretation of the Firthian concept of contextual meaning, according to which words enter "into meaningful relations with other words around them" (Sinclair 2004, p. 25) and do not remain "perpetually independent in their patterning" (Sinclair 2004, p. 30).

The secondary focus is placed on the immediate verbal phraseological environment of terms, the so-called *term-embedding collocations* (Biel 2014b, p. 180) or *collocations with a term* (Kjær 2007, p. 509), which are generally organised around the prototypical structure [N_{term}+V] or in the reverse order [V+N_{term}], and denote actions that are possible to undertake with the base noun, or in this case, with the nominal multi-word term. In other words, collocations with a verb encase the knowledge-carrying multi-word terms in a script, integrating the general domain-specific picture.

3. Corpus and methodology

This study is developed on the basis of a corpus of 92 judgments and decisions of the ECtHR in the English language, delivered on issues of bioethics, for the total of 1,130,548 words.

The corpus is subdivided into 10 thematic areas of interest ("subcorpora" or "micro-corpora", see Table 1), based on the categorisation adopted by the Council of Europe and the ECtHR (Research Report on Bioethics of the Council of Europe 2016, p. 4): physician-assisted suicide, consent to medical examination or treatment, ethical issues concerning HIV, retention of biodata (including DNA and fingerprint samples), medically assisted procreation, reproductive rights (including prenatal diagnosis and the right to a legal abortion), liability of health professionals, right to know one's biological identity, transgender issues and rights of children born out of surrogacy agreements.

Corpus	Texts	Words	Types
Physician-assisted suicide	10	112,822	5,675
Consent to medical examination or treatment	10	115,320	5,914
Ethical issues concerning HIV	13	162,452	7,595
DNA samples and retention of biodata	9	90,347	4,979
Medically assisted procreation	6	98,545	5,466
Reproductive rights	10	186,074	7,059
Liability of health professionals	6	67,960	4,608
Right to know one's biological identity	11	91,410	4,014
Transgender issues	14	156,736	6,142
Children born out of surrogacy	3	48,882	3,379
Total:	**92**	**1,130,548**	**17,118**

Table 1
Composition of the ECtHR case-law corpus.

The judgments and decisions included in the corpus were collected by the thematic criterion of bioethical relevance and by the linguistic criterion, as only texts in English were included. An important point should be made concerning the time span. The texts under analysis were produced in a period ranging from the 1990s (early transgender and HIV cases) until 2017. For the sake of categorisation, I roughly subdivide this time span into three periods: 1990-2000, 2001-2009, 2010-2017. The temporal placement of bioethically relevant ECtHR case-law has provided evidence of the increasing importance of this topic in the practice of this supranational court.

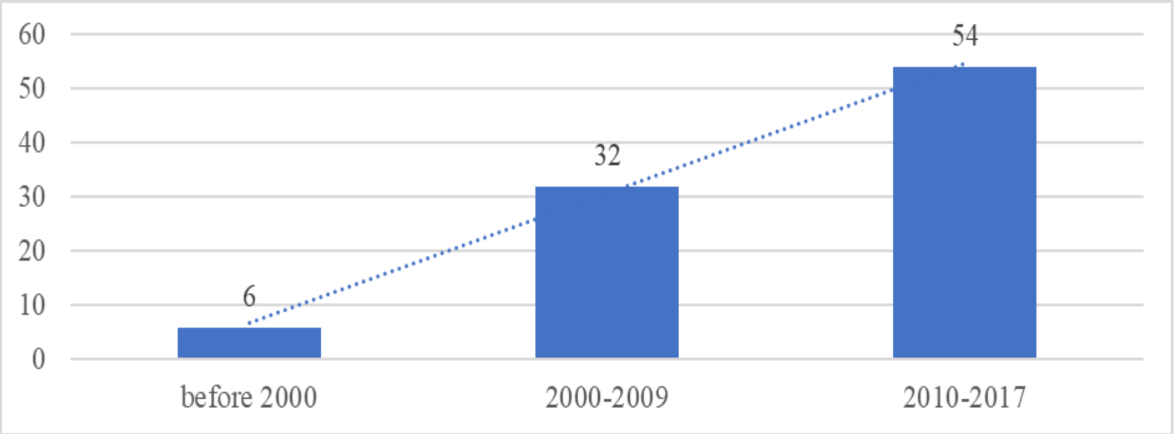


Figure 1
Number of bioethically relevant judgments and decisions in the ECtHR case-law.

As Figure 1 demonstrates, over the years the number of judgments and decisions concerning bioethical issues has significantly increased, with a clear upgoing trend. As of now, there is a substantial number of communicated and pending cases concerning such issues as DNA sampling, and “we can perhaps expect more applications concerning subjects such as gene therapy, stem cell research and cloning in the future” (Research Report on Bioethics of the Council of Europe 2016, p. 4).

The general methodological approach is quantiqualitative. I use corpus methodology to identify specialised phraseological units and further qualitatively analyse how specialised meanings are expressed and encoded through these patterns. In particular, this study looks at how bioethical patterns intertwine with the legal matrix of judgments and how this interaction creates the complex fabric of legal discourse on bioethics.

The research adopts *TermoStat 3.0* (Drouin 2012) software for multi-word candidate term extraction, based on statistical and linguistic criteria. The software was created for domain-specific information retrieval and uses an in-built reference corpus of non-specialised texts. Drouin explains (Drouin 2012) that the reference corpus is about 8,000,000 occurrences, which corresponds to 465,000 word types. The reference corpus consists of two parts: a) journal articles and b) a part of the British National Corpus (BNC). I also use *WordSmith Tools 6.0* (Scott 2015) software, and in particular its concord, patterns and clusters functions, and *AntConc 3.4.4* (Anthony 2014) to verify multi-word term validity as well as for lexical analysis and text search.

The selection of multi-word terms is primarily based on their specificity score as compared to TermoStat’s non-specialised reference corpus, which is calculated directly by the software. The software also allows us to order the multiword terms by log-likelihood or chi^2 score, and these scores have been consulted, too. As specificity score, originally proposed by Lafon (1980), focuses particularly on the specificity of lexical items, “based on the assumption that a technical corpus contains a set of lexical items that are closely related to its subject and subject area” (Drouin 2003, pp. 99-100), this parameter has been chosen as the primary criterion. In addition, I double checked my candidate terms using the frequency of a given multi-word term as a secondary criterion. Such an organisation allows me to focus only on those items that are highly specific to the analysed linguistic variety, without losing of sight the purely quantitative component of high recurrence.

Structurally, the multi-word terms and collocations with a term are categorised by their grammatical types, by the number of elements in the n-gram and by the thematic component. Next, using the concord, patterns and clusters functions of WordSmith Tools 6.0 I checked the phraseological environment of the candidates chosen for the analysis before proceeding with the classification.

4. Findings

4.1. Manual sorting and classification of multi-word terms

TermoStat Software has identified 4,026 potential multi-word terms of a nominal nature, ranging from 2-grams to 7-grams. I have selected only candidates with the specificity score above 5 (log likelihood above 20) and raw frequency of at least 5 occurrences; however, the latter criterion was of a secondary nature as my findings were sorted by specificity. I have manually checked the selected candidates verifying the software's precision with regard to a) the terminological status of a given candidate and b) the correct assignment of a grammatical category.

Concerning the terminological status, I draw on a number of classifications of legal terms (Mellinkoff 1963; Riley 1995; Alcaraz Varó, Hughes 2002; Chromà 2011) and include in the respective categories 1) purely technical terms (occurring only in a specific domain, such as "criminal offence" for legal discourse or "in vitro fertilisation" for bioethical discourse); 2) semi-technical or mixed terms (such as "dissenting opinion") or 3) recontextualised everyday vocabulary acquiring specialised meaning in the context (e.g. "fair balance").

Based on the above criteria, I have classified the sorted multi-word terms into a) legally-relevant (e.g. "domestic court"), b) bioethically-relevant (e.g. "gender reassignment"), c) of an ambiguous relevance (e.g. "family life") and d) general ("general interest"). The latter category was not treated as terminological. Ambiguous multi-word terms gather terms which could be used both in bioethically-relevant legal discourse and in less specialised texts (e.g. "state of health"), as well as terms which could be classified as both legally and bioethically-specific, e.g. "lawful abortion". However, the bioethical component is intended to be the prevailing one. As bioethics is a very elusive concept with blurred borders, the presence of two categories allows a certain flexibility in identifying different cases of potentially bioethically-relevant terms. In general, such a thematic division is possible on account of a multi-word nature of terms, which allows us to contextualise and assess their meaning. The thematic distinction is introduced as an approximate indicator and should not be interpreted in absolute terms.

The manual check corroborated the reliability of the software; however, I had to eliminate incomplete chunks of multiword expressions, such as "legal parent-child" ("legal parent-child relationship"). The software proved also to be reliable in reading syntax, as it made errors only with regard to Latin phrases (e.g. "mutatis mutandis", "de facto", "locus standi", etc), some verbal constructs (e.g. "give birth" or "claimed compensation" was software marked as [Adj + N], although the concordance proved it to be [V+N]).

It is interesting to note that legal multi-word terms represent more than a half of all sorted terms (see Table 2), whereas bioethically charged multi-word terms and ambiguous terms amount respectively to 23% and to 25%. However, if the latter two categories are combined, the proportion becomes 54% to 48%. Such a high proportion can be considered, at a preliminary level, as an indication that legal and bioethical terms in the ECtHR judgments overlap.

	Legal		**Ambiguous**		**Bioethical**		**Total**	
	Freq	%	Freq	%	Freq	%	Freq	%
	708	**54%**	**330**	**25%**	**285**	**23%**	**1,323**	**100%**
n-grams								
2-grams	517	73%	235	71%	216	76%	971	73%
3-grams	139	20%	64	19%	45	16%	247	19%
4-grams	46	7%	27	8%	21	7%	94	7%
5-grams	3	0.4%	2	1%	3	1%	9	1%
6-grams	1	0.1%	1	0.3%	-	-	2	0.2%

Table 2
Multi-word terms sorted by domain and by n-grams.

In terms of the number of elements in the cluster, bigrams are the most recurrent category with 73% of occurrences in total. 3-grams and 4-grams represent respectively 19% and 7%, and the other categories are significantly less frequent. The distribution of n-grams in different thematic categories reflects the general tendency, yet one can note that bioethical multi-word terms are most frequently bigrams (76%).

Concerning the morphological composition, the most productive pattern of a multi-word term, amounting to 56% of all multi-word terms analysed, is [Adj + N] (748 different multi-word types). This pattern is represented by such expressions as “private life”, “domestic court”, “positive obligation”, “medical treatment”, etc. In an inter-thematic perspective, it is the most recurrent pattern in legal multi-word terms (60%). Remarkably, the second and third most productive patterns differ across the thematic blocks: bioethical multi-word terms rely on [N+N] structures in 24% (“home birth”, “gender reassignment”, “insemination facility”), whereas legal multi-word terms employ it only in 11% of multi-word types (“remand prison”, “police custody”), which would suggest a higher nominalisation of bioethical discourse. At the same time [N + Prep + N] pattern amounts to 14% of legal multi-word terms (“margin of appreciation”, “declaration of incompatibility”) and only to 8% of bioethical terms (“removal of tissue”, “donation of embryos”). It would seem thus that information is more densely packed in bioethical multi-word terms, with less explicit prepositional links.

The fourth position with 3% of occurrences on the average is divided between the structures [Adj + Adj + N] (e.g. "joint dissenting opinion", "full legal recognition", "presumed biological father") and [Adj + N + N] ("intensive care unit", "civil status register"). Finally, 2% are represented by the [N + Prep + Adj + N] pattern (e.g. "lack of legal recognition", "exhaustion of domestic remedy", "use of human embryo").

	Legal		**Ambiguous**		**Bioethical**		**Total**	
Composition	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**
Adj + N	422	60%	176	53%	148	52%	748	56%
N + N	79	11%	58	17%	67	24%	205	15%
N + Prep + N	97	14%	38	11%	22	8%	150	11%
Adj + Adj + N	22	3%	10	3%	7	2%	39	3%
Adj + N + N	15	2%	11	3%	10	4%	36	3%
N + Prep + Adj + N	16	2%	10	3%	7	2%	33	2%
Adj + CConj + Adj + N	19	3%	10	3%	4	1%	31	2%
Latin	13	1.8%	-	-	3	1%	16	1%
N + N + N	3	0.4%	8	2%	3	1%	15	1%
Adj + N + Prep + N	6	1%	4	1%	2	1%	12	1%
V + N (NomPhr)	4	0.6%	1	0.3%	2	1%	12	1%
N + Prep + N + N	1	0.1%	3	0.9%	3	1%	7	0.5%
N + Adj + N	1	0.1%	1	0.3%	1	0.4%	3	0.2%
Adj + N + Prep + Adj + N	1	0.1%	-	-	1	0.4%	3	0.2%
Adj + N + N + N	1	0.1%	1	0.3%	1	0.4%	3	0.2%
Other patterns	8	-	-		4	-	20	2%

Table 3
Multi-word terms sorted by grammatical composition.

4.2. General collocational environment and combinability of multi-word terms

During the sorting I came across some economic terms ("marginal lending rate", "default interest"). Interestingly, these economic patterns were embedded into large chunks of a highly recurrent nature together with legal phraseological units.

(1) The Court considers it appropriate that the *default interest rate* should be based on the *marginal lending rate* of the European Central Bank, to which should be added *three percentage points*.[1]

(2) that from the expiry of the above-mentioned three months until settlement *simple interest* shall be payable on the above amounts at a rate equal to the *marginal lending rate* of the European Central Bank during the *default period* plus three *percentage points*.

[1] Emphasis is added in all examples.

Example (1) is a 31-gram occurring in 16 texts, and example (2) is a 42-gram repeated in 30 texts. Both larger formulae feature in the judgment's disposition part concerning equal satisfaction (compensation money). Their size and high recurrence make it possible to define them as routine formulae, structural building blocks or text-organising patterns (Biel 2014b: 178) of the ECtHR judgments.

Analysis of the phraseological environment of legal multi-word terms casts light on their combinability and embeddedness in larger phrases. These larger units are frequently quotations of the European Convention on Human Rights. For instance, example (3) is 45-gram used in 21 texts and it is an excerpt from Article 14. Example (4), based on Article 8.2 of the Convention, is the longest and the most recurrent lexical bundle in all texts analysed: it counts 71 grams and is used in 62 texts.

(3) The *enjoyment of the rights and freedoms* set forth in [the] Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, *political or other opinion*, *national or social origin*, association with a national minority, property, birth or other status.

(4) There shall be no interference by a *public authority* with the *exercise of this right* except such as is in accordance with the law and is necessary in a *democratic society* in the *interests of national security*, *public safety* or the *economic well-being of the country*, for the *prevention of disorder or crime*, for the *protection of health or morals*, or for *the protection of the rights and freedoms* of others.

Bioethical multi-word terms do not exhibit a similar tendency to be grouped in extremely large and repetitive text chunks, although ambiguous terms of a twofold relevance can figure in longer clusters as they are cited in the case-law:

(5) The Court reiterates/ notes that the notion of *"private life"* within the meaning of Article 8 of the Convention is a broad concept which encompasses, inter alia, the right to …

An interesting observation emerged concerning the placement of bioethically relevant multi-word terms. In all the judgments analysed such terms and notions were extensively addressed and discussed in the "Relevant domestic law" or "Relevant domestic and international law and practice" part of the judgment.

(6) According to Article 1592 of the Civil Code, *a child's father is either the man who on the date of the child's birth was married to the child's mother* (no. 1), or *the man who acknowledged paternity* (no. 2), or *whose paternity is judicially established under Article 1600d of the Civil Code* (no. 3). [Case of Ahrens v. Germany]

(7) Section 55 of this law *defines gestational surrogacy as the fact of bearing and handing over a child on the basis of a contract concluded between the surrogate*

mother and the intended parents, who provide their own genetic material. […] Decree no. 107, issued by the Minister of Health on 30 August 2012, *defines gestational surrogacy as a contract entered into between the surrogate mother and the intended parents who have used their genetic material for the conception*. [Case of Paradiso and Campanelli v. Italy]

(8) In order to remedy *reproductive problems arising as a result of human sterility or infertility*, recourse may be had to medically assisted reproduction in the conditions and in accordance with the procedures provided for by this Law, which guarantees the rights of all the persons concerned, including those of the subject thus conceived. [Case of Parrillo v. Italy]

Naturally, these patterns could not achieve a high number of repetitions on account of thematic variation across the subcorpora. A general – and somewhat surprising – tendency is however clear: the highest concentration of bioethically charged terms is found in the section devised to deal with a legal overview of a certain national system. The rationale standing behind this placement peculiarity could lie in a combination of factors: a) legal discourse strives towards precision and relevant terms are frequently duly defined to avoid ambiguity; b) bioethical concepts, being external to the average legal situation, need explanation as they are entextualised from another scientific reality, i.e. their meaning is extracted from one discourse and consequently inserted into another through a process of de-contextualization and re-contextualisation in another context (Silverstein, Urban 1996, p. 15 in Garzone 2014, p. 79). In general, this placement allows us to fix artificially the meaning of bioethical and ambiguous terms in the context of judgments, and bioethical terms invariably collocate with legal terms and expressions in these definitions and elsewhere, proving a solid link between these two discursive realities.

The sorting revealed a number of collocations of the [Adj + N] pattern, with a relatively low degree of terminologicality, such as "present case" / "instant case", most often used within a nominal phrase "in the present / instant case", or qualifications "second applicant", third and fourth applicants", based on the number of applicants involved in a case. These and similar occurrences were marked as lexical collocations following Biel (2014b, p. 181) as they merely perform a referential function by using a qualifying adjective in adjacency to a legal term. It must be noted that the distinction and separation of terms from collocations is arbitrary to a certain degree (see, for instance, Heid 2001, p. 791). I identified also a number of collocations from the general language use ("general interest", "last-mentioned date", "chilling effect"), which were inserted into nominal phrases or qualifications (Bhatia 2006, p. 2) and, although not being terminological, increased the legal flavour as they served the purpose of creating the phraseological matrix of texts.

The [N + N] and [N + Prep + N] structures also frequently co-occurred with qualifying adjectives. For instance, the most recurrent legal multi-word term of the latter pattern – "margin of appreciation" (frequency: 551; specificity: 54,68) – co-occurred with such adjectives as "wide" (102 hits) "broad" (20), "narrow" (20), "significant" (3) and "relevant" (3). Another combinatorial tendency observed concerned a number of pseudo-multinominal terms, composed of "the issue of" + N, "the concept of" + N or "the notion of" + N. In these bundles the first element was somewhat redundant as it did not carry an additional terminological meaning, but mainly enhanced the nominalised nature of judgments.

The manual stage of sorting, although highly time-consuming, has also allowed me to trace a certain pattern with respect to the collocational environment of nominal multi-word terms. While I had to eliminate such chunks as "respect of non-pecuniary damage" (truncated "in respect of non-pecuniary damage") and similar structures, I noticed that multi-word terms – legal, bioethical and ambiguous – were frequently co-occurring with complex prepositions, built around the pattern [Prep + (Art) + N + Prep]. Unfortunately, TermoStat does not retrieve the above structures as it is programmed to trace terminological choices and to disregard functional vocabulary. Consequently, complex prepositions should be addressed in a separate paper, using a different software. However, using WordSmith Tools 6.0 I briefly looked at 4-grams – uninterrupted sequences of four elements – and identified 416 different prepositional phrases, 60 of which (about 15%) were highly recurrent and frequent complex prepositions used with both legal and bioethical multi-word terms, such as *within the meaning of* (293); *on the basis of* (284); *as a result of* (233); *in the context of* (230); *for the purposes of* (217). 3-element complex prepositions were even more frequent and included such items as *in accordance with* (607), *in respect of* (538), highlighting again a high prominence of these grammatical structures for the texts under analysis.

4.3. Term-embedding collocations with a verb

Apart from a general collocational environment of multi-word terms, I looked at a special category of collocations, the so-called term-embedding collocations (Biel 2014b, p. 180). These collocations "form the skeleton of legal rules by providing action and enabling terms to enter into relations" (Biel 2014b, p. 180). In other words, these are

> Collocates of terms which embed terms in cognitive scripts and the text, evidencing combinatory properties of terms. […] They establish links between terms and elements of conceptual frames. […] Subtype-denoting collocations are often subject to terminologisation and form distinct terms. (Biel 2014b, p. 180)

Prototypically these phraseological units are built around the structure [N_{term}+V] or in the reverse order [V+N_{term}], and denote actions that are possible to undertake with the base noun.

Having sorted the term-embedding collocations by the association score, based on the log-likelihood test, I have attributed to the selected candidates the same classification already elaborated for multi-word terms: legal, bioethical, ambiguous, as well as general, and have analysed those items that had an association score above 25.

Thematic class	n. types	%	Examples	Inclusion in the phrase
Bioethical	137	19%	Commit suicide	helping someone to commit suicide is not a punishable offence
Legal	381	52%	Enjoy margin of appreciation	states enjoy a wide margin of appreciation
Ambiguous	112	15%	Found family	right to marry and to found a family
General	108	15%	Cast doubt	Genetic evidence cannot cast doubt
Total	738	100%		

Table 4
Term-embedding collocations sorted by the thematic element.

As expected – based on the findings of multi-word terms – term-embedding collocations with a legal thematic element prevail, with 52% of different collocations belonging to this category. They are followed by bioethically relevant collocations with a verb (19%). 15% of cases are ambiguous and 15% of collocations with a verb can be classified as general in that they can be used in general context. These results are to be treated with caution. Some verbs can carry a legal thematic connotation (“exhaust”, “dismiss”, see (9)) or a bioethical connotation (“conceive”, see (10)); however, in most cases it is difficult to associate any bioethical connotation just to a verb, also in light of the elusive nature of the very concept of bioethics and its highly interdisciplinary character. For instance, in (11) “order surveillance” has a distinct legal flavour, although in the general context of the utterance it acquires also a bioethical relevance. Consequently, the above subdivision is to be intended as approximative.

(9) The Government’s objection as to the failure by the first and second applicants to *exhaust domestic remedies* must therefore be *dismissed*. [Case of I.G. and Others v. Slovakia]

(10) He observed that an agreement whereby a woman undertook to *conceive and bear a child* and <u>relinquish it at birth</u> was null and void in accordance with the public-policy principle that the human body and civil status are inalienable. [Case of Mennesson v. France]

(11) To sum up, it should be concluded that *ordering surveillance* of insured persons by accident insurers in the context outlined in point 4.3 [above] is permitted; the results

of such surveillance can thus in principle be used for assessment of the issues in question .... [Case of Vukota-Bojic v. Switzerland]

It is crucial to look at the phrase inclusion, as often collocations classified as general in per se are included in a context, where they assume bioethical or legal relevance. For instance, "women travel" is classified as general. However, if we look at its inclusion in a sentence, we find that the utterance undoubtedly belongs to the bioethical field, which proves the importance of the phraseological perspective to assess the specificity of this discourse.

(12) The Government also underlined that the impugned restrictions had led to a significant reduction in Irish *women travelling* to the United Kingdom *for an abortion* (6,673 women in 2001 travelled and 4,686 women did so in 2007) and to one of the lowest levels of maternal deaths in the European Union. [Case of A, B and C v. Ireland]

Interestingly, in most term-embedding collocations the nominal element acts as a grammatical object (56%, see Table 5); however, it could have been expected as I include in this category direct objects ("exhaust domestic remedies"), indirect objects ("award the applicant (just satisfaction)") and prepositional objects ("comply with a requirement"). Unfortunately, the software does not distinguish between a prepositional object and a prepositional complement (Quirk *et al*. 1985, p. 1201), consequently I had to manually reclassify some of the phrases as collocations with a prepositional complement (18%, e.g. "travel for abortion"), which are a borderline case between term-embedding collocations and lexical collocations. There are 28% of cases where the nominal element acts as a grammatical subject ("Court reiterates"). When the nominal element is the institutionalised actor involved in the proceedings (e.g. Court or applicants), it frequently collocates with verbs with communicative meaning ("the applicants claim") or semantics of mental activity ("the Court considers"). Where both subject and object functions are traced, e.g. "institute proceedings" and "proceedings are instituted", both are counted separately.

	Legal		**Ambiguous**		**Bioethical**		**Total**	
Function of the N	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**
Object	205	54%	78	70%	86	63%	414	56%
Subject	108	28%	28	25%	31	23%	213	29%
Complement	68	18%	6	5%	20	15%	111	15%
Total	381		112		137		738	

Table 5
Term-embedding collocations sorted by the function of the nominal element across various thematic groups.

Table 6 gathers 25 first-ranking term-embedding collocations sorted by their association score (log-likelihood, see column 4 in Table 6), i.e. by their domain specificity, and examples of the collocations being used in the corpus. Interestingly, some collocations with a verb, although following the general [V + N] pattern, differ from others as they seem to demonstrate a higher degree of fixedness and indivisibility (Hudson 1998, pp. 5-10, 35). Even though such collocations as "fall within the ambit of" (association score: 150; frequency: 55) or "fall within the scope of" (association score: 423; frequency: 59) contain the nouns "ambit" and "scope", the former is not used outside of these expressions, and the latter is prevalently used as "within the scope of" collocating also with "come", "bring" and "take", demonstrating a highly convergent pattern. It would seem that the pattern "within the scope/ambit of" exhibits a high degree of grammaticalization and functions almost as a complex preposition.

Verb	Noun	Fre	AssSc	Example
exhaust	remedy	72	692	the applicant had not exhausted domestic remedies
commit	suicide	104	623	commit suicide
strike	fair balance	77	617	to strike a fair balance
bear	child	131	595	child (was) born
practise	lawyer	63	568	lawyers practising in Riga
originate	application	78	527	the case originated in an application
annex	judgment	49	480	opinions are annexed to this judgment
adopt	date	74	478	judgment adopted on that date
bear	wedlock	68	478	born out of wedlock
enjoy	margin of appreciation	68	445	states enjoy a wide margin of appreciation
add	percentage points	43	444	to which should be added three percentage points
undergo	gender reassignment	81	435	she had undergone gender reassignment surgery
fall	scope	59	423	fall within the scope of
afford	just satisfaction	53	392	the court shall, if necessary, afford just satisfaction
balance	competing interests	51	374	to balance competing interests
found	family	43	356	right to marry and to found a family
originate	case	71	349	the case originated in an application
incur	costs and expenses	40	340	costs and expenses incurred before the court
institute	proceedings	60	329	department had failed to institute proceedings
terminate	pregnancy	41	310	the applicant's pregnancy was not terminated
guarante e	convention	67	272	the rights guaranteed by the convention
impose	obligation	48	267	the Convention did not impose an obligation on contracting states
challenge	paternity	43	262	allowing the biological father to challenge paternity without restrictions
travel	abortion	40	261	women travelling abroad for abortion

Table 6
25 most domain-specific term-embedding collocations
organised in a decreasing order according to the association score.

As I already argued elsewhere (Nikitina 2018, p. 200), adapting and slightly modifying Biel's (2014a, 2014b) classification, complex prepositions act as grammatical patterns on account of their prefabricated nature and grammatical functions as well as functional distinctiveness for legal discourse in general. It emerges that such complex phrases as "within the scope of", "within the ambit of" and "within the sphere of", which could be rephrased by "under", collocate with "fall" on the left or other verbs with a reduced semantic charge, making the whole expression functionally prefabricated and classifiable as a borderline case between term-embedding collocations and grammatical patterns.

The analysis sheds light on a high concentration of specialised deverbal nouns in term-embedding collocations, which reduced the recurrence of verbs. For instance, instead of using a more general verb "to treat" / "be treated", the corpus shows a number of verbs collocating with "treatment" as an object (with association score indicated in brackets): *receive treatment* (253), *undergo treatment* (212), *accept treatment* (114), *refuse treatment* (160) as well as *subject to treatment* (92), *administer treatment* (52), *discontinue treatment* (38), *propose treatment* (93), *proscribe treatment* (129), *start treatment* (34) and *withdraw treatment* (226), some of which could have been replaced by "treat". "Treat" is used in term-embedding collocations with a subject: *doctor treats* (139), and in collocations with a direct object: *treat a patient* (121), *treat a child* (82), *treat a woman* (72) and *treat infection* (39). All of the nouns co-occurring with "treat" could collocate as indirect or prepositional objects with "treatment". Similarly, the ambiguous term – both bioethically and legally relevant – "consent" is not used as a verb, but only as a noun in term-embedding collocations *withdraw consent* (205), *withhold consent* (134), *obtain consent* (70) or *express consent* (55). These verbo-nominal constructs demonstrate high association scores and confirm their domain specificity, marking legal discourse on bioethics as tendentially nominalised discourse.

Moreover, the analysis identified a number of collocations belonging to the category of *verbo-nominal phrases*, or *verbonominants*, i.e. those [V + N^{obj}] structures, where the noun – usually a deverbal noun - carries the lexical meaning of the whole pattern, and the verb is "semantically reduced or light" (Leech *et al*. 2009, p. 166), e.g. "reach a conclusion" instead of "to conclude". This phenomenon "at the borderline between syntax and lexicon" (Leech et al. 2009, p. 166) could be effectively addressed through the paradigm of phraseology. These patterns, known under a great variety of labels in the academic literature[2] increase the nominal character of discourse,

[2] See Leech *et al*. 2009, pp. 166-167 for an overview.

closely associated with legal (cf. Mellinkoff 1963; Crystal, Davy 1973; Tiersma 1999) and generally specialised discourse (Leech *et al*. 2009). In legal discourse this phenomenon is described as "preference for nouns and nominalisations (nouns derived from verbs, such as 'consideration' or 'injury') over verbs ('consider' or 'injure')" (Tiersma 2015 [2006], p. 28). Verbo-nominal constructions are generally composed of verbs of common meaning, such as "take", "make", "give", "have" or "put" (Claridge 2000, p. 71), which do not behave as domain-specific, and thus are not retrieved by the log-likelihood specificity test. The software has retrieved mostly cases of domain-specific verbonominants, such as those collocating with "reach": *reach agreement*, *reach conclusion* and *reach decision*. While these constructs could be rephrased by a single verb, it is a clearly dispreferred tendency.

5. Conclusions

This article was designed following a general descriptive goal of presenting some linguistic peculiarities of the phraseological organisation of the ECtHR judgments dealing with a variety of bioethical issues. The analysis of term-related patterns, their structure and thematic components confirms that the skeleton of the judgments, with 54% of multi-word terms and 52% of term-embedding collocations with a verb, is composed of phraseological units with a legal thematic component. The smaller multi-word units are often packed into lengthy prefabricated chunks of texts, up to 71-grams, and recur in different judgments. These larger units mark different structural blocks of the judgment: the beginning, transitions between different parts, the disposition and the end, and contribute to the general perception of standardisation and formulaic character of this genre.

Apart from the structural matrix of the judgment, which remains purely legal, it appears possible to talk about a cross-fertilisation between legal and bioethical areas. In general, the findings demonstrate that bioethical and legal notions are intertwined in a complex relationship, up to the point where it becomes problematic to distinguish between the two elements, or where either of the meanings is possible, or where words from general language acquire specialised meaning in the context. It emerges that bioethical notions are entextualised into the legal matrix of judgments without altering the generic structure of the latter. While bioethical term-related patterns are generally disseminated throughout the judgments, they are especially frequent in the parts describing facts and national and international points of law. The latter concentration – typically under the form of definitions and explaining provisions – could be construed as a sign of legal precision on the one hand and as an instance of specialised knowledge transfer from the scientific domain and its incorporation into the legal domain, which could be interpreted as a strategy of knowledge management of the participants (van Dijk 2003).

The structural analysis of multi-word terms allowed to trace a slight difference in knowledge presentation between legal and bioethical units. It emerged that, among different thematic classes, bioethical multi-word terms use the [N + N] pattern most frequently, and disprefer a more explicit prepositional structure of the [N + Prep + N] type. The tendency is reversed in respect of legal multi-word terms. These preferential morphological patterns translate into a higher degree of nominalisation of bioethical discourse compared with legal discourse. The nominalisation is further exacerbated by the recurrence of the so-called verbo-nominal structures, i.e. the preference to use a semantically reduced verb with a deverbal noun/ nominal expression. The collocational environment of term-related units includes combinations between various multi-word terms, frequently achieved through the use of complex prepositions or grammaticalized verbal constructs. It emerges, however, that bioethical multi-word terms tend to appear in close proximity to legal multi-word terms, or verbal constructs with a legal thematic element.

Analysis of multi-word terms and their immediate collocational environment proved to be a valid starting point to assess the entextualisation processes, demonstrating some convergent and divergent tendencies, both in the composition and in their phraseological behaviour. The results of this research and linguistic descriptions produced could and will be employed as a starting point to analyse further mechanisms of legal and bioethical knowledge dissemination, during its migration from an institutionalised setting to a more liberal web space under the form of blogs.

Bionote: Jekaterina Nikitina has a PhD degree from the University of Milan in “Linguistic, Literary and Intercultural Studies in European and Extra-European Perspectives”. Her main academic interest lies in the field of Legal Discourse and Legal Translation Studies, with her PhD thesis focusing on the analysis of phraseological patterns in written pleadings before the European Court of Human Rights translated into English from Russian and from Italian. Jekaterina’s areas of interests extend to LSP theories and applications, knowledge dissemination dynamics, discourse of healthcare, medicine and bioethics, applying qualitative and quantitative, specifically corpus linguistics, analytical approaches. She is an external collaborator of the research unit of Milan in the PRIN project dealing with knowledge dissemination across media in English. Her published academic work includes publications on legal translation and dissemination of bioethical knowledge, with a focus on medically assisted procreation and gene editing.

Author’s address: jekaterina.nikitina@unimi.it

References

Alcaraz Varò E. and Hughes B. 2002, *Legal Translation Explained*, St. Jerome Publishing, Manchester.

Anthony L. 2014, *AntConc* (Version 3.4.4) [Computer Software], Waseda University, Tokyo. http://www.laurenceanthony.net/.

Bergenholtz H. and Tarp S. 1995, *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*, Benjamins, Amsterdam/Philadelphia.

Bhatia V. 2006, *Legal Genres*, in Brown K. (ed.) *Encyclopedia of Languages and Linguistics*, Elsevier, Oxford, pp. 1-7.

Bhatia V. 2014 [2004], *Worlds of Written Discourse: A Genre-Based View*, Continuum, London.

Biel Ł. 2014a, *Lost in the Eurofog. The Textual Fit of Translated Law*, Peter Lang, Frankfurt am Main.

Biel Ł. 2014b, *Phraseology in Legal Translation: a Corpus-based Analysis of Textual Mapping in EU Law*, in Cheng L., Kui Sin K. and Wagner A. (eds), *Ashgate Handbook of Legal Translation*, Ashgate, Aldershot, pp. 177-192.

Bioethics and the Case-law of the Court 2016 [2009], Research Report, Council of Europe / European Court of Human Rights. http://www.echr.coe.int/Documents/Research_report_bioethics_ENG.pdf (27.06.2017).

Calsamiglia H. and van Dijk T. 2004, *Popularization Discourse and Knowledge about Genome*, in "Discourse and Society" 15 [4], pp. 369-389.

Chromà M. 2011, *Synonymy and Polysemy in Legal Terminology and Their Applications to Bilingual and Bijural Translation*, in *Legal Terminology: Approaches and Applications*, *a* special issue of "Research in Language" 9 [1], pp. 31-50.

Claridge C. 2000, *Multi-word Verbs in Early Modern English: A Corpus-Based Study*, Language and Computers 32, Rodopi Publishers, Amsterdam.

Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine, European Treaty Series - No. 164, Oviedo, 4.IV.1997. https://rm.coe.int/168007cf98 (27.06.2017).

Crystal D. and Davy D. 1973, *The Language of Legal Documents*, in Crystal D. and Davy D. (eds.), *Investigating English Style*, Indiana University Press, Bloomington, pp. 193-217.

Drouin P. 2003, *Term Extraction Using Non-Technical Corpora as a Point of Leverage*, in "Terminology" 9 [1], pp. 99-117.

Drouin P. 2012, *TermoStat Web 3.0* [software], Université de Montréal. http://termostat.ling.umontreal.ca/doc_termostat/doc_termostat_en.html

ECtHR *Rules of Court* as of 14 November 2016. https://www.echr.coe.int/Documents/Rules_Court_ENG.pdf (26.02.2018).

Garzone G. 2014, *News Production and Scientific Knowledge: Exploring Popularization as a Process*, in Bongo G. and Caliendo G. (eds.), *The Language of Popularization: Die Sprache der Popularisierung*, Peter Lang, Bern, pp. 73-107.

Granger S. and Paquot M. 2008, *Disentangling the Phraseological Web*, in Granger S. and Meunier F. (eds.), *Phraseology: An Interdisciplinary Perspective*, Benjamins, Amsterdam/Philadelphia, pp. 27-49.

Heid U. 2001, *Collocations in Sublanguage Texts: Extraction from Corpora*, in Wright S.E. and Budin G. (eds.), *Handbook of Terminology Management. Application-*

Oriented Terminology Management, vol. 2, Benjamins, Amsterdam-Philadelphia, pp. 788-808.
Hudson J. 1998, *Perspectives on Fixedness: Applied and Theoretical*, Lund University Press, Lund.
Hyland K. 2008, *As Can Be Seen: Lexical Bundles and Disciplinary Variation*, in “English for Specific Purposes” 27, pp. 4-21.
Kjær A.L. 1990, *Context-Conditioned Word Combinations in Legal Language*, in “Terminology Science & Research, Journal of the International Institute of Terminology Research – IITF” 1 [1-2], pp. 21-32.
Kjær A.L. 2007, *Phrasemes in Legal Texts*, in Burger H., Dobrovol’skij D., Kühn P. and Norrick N. (eds.), *Phraseology: An International Handbook of Contemporary Research*, Mouton de Gruyter, Berlin, pp. 506-516.
Lafon P. 1980, *Sur la variabilité de la fréquence des formes dans un corpus*, in “MOTS” 1, pp. 128-165.
Leech G., Hundt M., Mair Ch. and Smith N. 2009, *Change in Contemporary English: A Grammatical Study*, Cambridge University Press, Cambridge.
Mellinkoff D. 1963, *The Language of the Law*, Little Brown & Co, Boston.
Meyer I. and Mackintosh K. 1994, *Phraseme Analysis and Concept Analysis: Exploring a Symbiotic Relationship in the Specialized Lexicon*, in Martin W. (ed.), *Euralex 94 Proceedings*, Amsterdam, pp. 339-348.
Myers G. 2003, *Discourse Studies of Scientific Popularization: Questioning the Boundaries*, in “Discourse Studies” 5 [2], pp. 265-279.
Nikitina J. 2018, *Legal Style Markers and Their Translation: Written Pleadings at the ECtHR*, Edizioni accademiche italiane / OmniScriptum Publishing, Beau Bassin-Riga.
Riley A. 1995, *The Meaning of Words in English Legal Texts: Mastering the Vocabulary of the Law – A Legal Task*, in “Law Teacher” 29-30, pp. 68-83.
Quirk R., Greenbaum S., Leech G. and Svartvik J. 1985, *A Comprehensive Grammar of the English Language*, Longman, London.
Sager J. 1998, *Terminology: Theory*, in Baker M. (ed.), *Routledge Encyclopedia of Translation Studies*, Routledge, London-New York, pp. 258-262.
Scarpa F., Peruzzo K. and Pontrandolfo G. 2014, *Methodological, Terminological and Phraseological Challenges in the Translation into English of the Italian Code of Criminal Procedure*, in Gialuz M., Lupária L. and Scarpa F. (eds.), *The Italian Code of Criminal Procedure. Critical Essays and English Translation*, Wolters Kluwer / CEDAM, Padova, pp. 53-80.
Scott M. 2015, *Wordsmith Tools version 6* [computer software], Lexical Analysis Software, Stroud.
Sinclair J. 2004, *Trust the Text: Language, Corpus and Discourse*, Routledge, London/New York.
Teubert W. 2002, *The Role of Parallel Corpora in Translation and Multilingual Lexicography*, in Altenberg B. and Granger S. (eds.), *Lexis in Contrast*, Benjamins, Amsterdam/Philadelphia, pp. 189-214.
Tiersma P. 1999, *Legal English*, Chicago University Press, Chicago/London.
Tiersma P. 2015 [2006], *Some Myths about Legal Language*, in Solan L., Ainsworth J. and Shuy R. (eds.), *Speaking of Language and Law. Conversations on the Work of Peter Tiersma*, Oxford University Press, Oxford, pp. 27-34.
Universal Declaration on Bioethics and Human Rights, adopted by UNESCO's General Conference on 19 October 2005. http://unesdoc.unesco.org/images/0014/001461/146180E.pdf, (27.06.2017).

van Dijk T.A. 2003, *The Discourse-Knowledge Interface*, in Wodak R. and Weiss G. (eds.), *Critical Discourse Analysis. Theory and Interdisciplinarity*, Palgrave, London, pp. 85-109.

Widdowson H. 1979, *Explorations in applied linguistics*, Oxford University Press, Oxford.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 291-314
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p291
http://siba-ese.unisalento.it, © 2019 Università del Salento

A COGNITIVE, SOCIO-SEMIOTIC, LINGUISTIC, AND DISCURSIVE APPROACH TO POPULARISATION STRATEGIES IN INFOGRAPHICS

SONIA PIOTTI, AMANDA CLARE MURPHY[1]
UNIVERSITÀ CATTOLICA DEL SACRO CUORE

Abstract – Information graphics or infographics are multimodal discursive spaces created by the combination of data and information visualisation, typography and colour. As effective forms of information communication and popularisation, infographics are frequently used by international organisations and government bodies as a means of popularising complex topics linked to health, food safety, politics, business and the environment. In this paper, Ciuccarelli's 2012 concept of the visual macroscope is adopted as an interpretative lens on a small corpus of infographics from the World Health Organisation, together with the tools of socio-semiotic, linguistic and discursive analysis, applied in a bottom-up approach. From a socio-semiotic perspective, it is seen that layout, pictorials, colour, typography and the order of information combine to make the *Ideal* and the *Real* (Kress, van Leeuwen 1996) stand out; from a linguistic and discursive point of view, thematic organisation interacts with pictorial organisation to make salient information emerge; lexical repetition, unmarked declaratives, and constant theme enact the strategy of explanation, frequently used in popularisation discourse.

Keywords: infographics; popularisation; interdiscursivity.

1. Introduction

To adopt a term coined by George Miller (1983), one of the founders of cognitive psychology, all higher organisms are *informavores*, wanting to be informed in order to make better decisions. In a society where the amount and complexity of information continues to grow, infographics are one of the most appropriate tools for dealing with and communicating that information.

[1] The paper was thought out, developed and written by both authors, who share the authorship equally. In particular, Sonia Piotti is responsible for Section 3, 5 and 6, and Amanda Clare Murphy for Section 1, 2 and 4.

Infographics are a popular genre used to produce and disseminate information across a variety of contexts from a real-world and socio-cognitive perspective. They are effective tools in business and professional teaching programmes (Toth 2013), in education and in academic contexts (IARE 2013). Businesses and news organisations including the *New York Times*, the *Washington Post* and *The Guardian* regularly incorporate graphics into their written stories. Politicians, activists and television reporters use graphics and visualisations as a backdrop for stories about socially relevant issues global health, economics and election results. Many national and international health associations also use infographics to engage what is often a heterogeneous audience. The trend can be found in science magazines and scientific journals such as *The British Medical Journal*, *Weird* and *Nature*, which use infographics to provide their expert audiences with accurate, research-based information. In the preparation of this paper, we have also found that many international institutions adopt this type of communication tool. These include the World Health Organisation, the European Parliament and Commission, the Directorates for General Food and Safety, Migration & Home Affairs, and the Canadian Centre for Policy Alternatives, all of which use infographics for fast knowledge transfer between evidence producers, particularly scientific communities, and policy makers.

In these contexts, infographics can be regarded as a class of texts or communicative events that fall under the broader category of popularisation, defined by Calsamiglia and Van Dijk (2004, p. 370) as

> the transformation of specialized knowledge into ‘everyday’ or ‘lay’ knowledge, as well as a recontextualization of scientific discourse, for instance, in the realm of the public discourses of the mass media or other institutions.

In this paper we concentrate on analysing how specialised knowledge in infographics is communicated also by means of socio-semiotic, linguistic and discursive resources, fields hitherto unexplored in the literature.

The paper is divided into five sections. Section 1 provides definitions of infographics and outlines some trends in infographics research, suggesting some distinguishing marks of infographics as opposed to other multimodal genres. Section 2 contextualises infographics from a historical perspective, while Section 3 describes infographics as a genre from the perspective of graphic design, the only perspective available in the literature thus far. Section 4 delves into the corpus analysis, focussing on the socio-semiotic level first and then moving to the linguistic and discursive. Section 5 summarises the findings and discusses future research.

We wish to make a methodological premise, regarding the language data which is the object of the analysis. We refer to the data as a small specialised corpus. Use of this term derives from the conviction that there is no consensus among scholars about the optimal size of a corpus. In the preface to Ghadessy, Henry and Roseberry's (2001) book devoted to studies with small corpora, one of the fathers of Corpus Linguistics, John Sinclair states that

> a small corpus is seen as a body of relevant and reliable evidence, and is either small enough to be analysed manually, or is processed by the computer in a preliminary fashion […]; thereafter the evidence is interpreted by the scholar directly. (Sinclair 2001, p. xi)

It is important to note that the claims made about specialised corpora are different from those made about other corpora: as Ghadessy, Henry and Roseberry (2001, p. xx) point out, the investigation of genres or types of texts "is an area not well suited to the use of large corpora, in which a number of very different genres are typically mixed together".

As the purpose of this study is to investigate the distinguishing marks of infographics as new discursive spaces for popularisation of specialised knowledge, which we consider a genre, our conclusion from the above considerations is that there is justification for seeing the language data in the present study as a small specialised corpus.

2. Definitions of *infographic*

The term *Infographic* is a blend of *information graphic*, the graphic representation or visualisation of data and information (Rendgen 2012, p. 9). In this capacity, information graphics lie at the crossroads of disciplines as diverse as information science and architecture, information visualisation, data visualisation, graphic design, communication and media studies and history. Different approaches to infographics can be discerned across these disciplines. Information science and visualisation are mainly interested in the process of the visual conversion of data and information as a significant part of information architecture. In traditional research activities into information visualisation, in fact, the term *visualisations* is preferred to *infographics* to refer to tools that may incorporate a variety of media, including text, images and video (Segel, Heer 2010).

We will now consider some definitions of infographics, mainly coming from the field of graphic design and art history. Infographics are defined as multimodal: for Krum (2014, p. 6), a graphic designer, an *infographic* is specifically "a larger graphic design that combines data visualizations, illustrations, text, and images together […]". Rendgen (2012, p. 9), an art historian with an interest in visual culture and technology, defines them as "hybrids and hence difficult to define", as they consist of "text, image and geometric shapes […]". A distinguishing mark highlighted by Smiciklas (2012), a graphic designer, is their property of being self-contained texts, the different modes being indissolubly linked together into a single entity, which "does not follow automatically from the data, but has to be developed [...]" (Rendgen 2012, p. 9).

A fourth characteristic pointed out by Krum (2014, p. 6) is that good infographic design is the result of the process of *storytelling*, i.e. an account of a series of events structured and presented in a tightly controlled progression, (*Oxford English Dictionary*, s.v. *storytelling*) and brought into a context that people can understand, remember, discuss and share. *Storytelling* was first identified as a key element of information design and visualisation by Gershon and Page (2001);[2] the idea was taken further by Segel and Heer (2010) with the coining of the term *narrative visualisation* (see also Lankow *et al*, 2012). These scholars developed a framework suggesting that use of visual narrative tactics and visual rhetorical devices guide the reader's attention and thus minimise cognitive cost, especially in journalistic storytelling. Drawing on news media data visualisation, they identified three visual *narrative* tactics: *highlighting* (through colour, motion, framing, size, audio, and more) *transition* and *guidance*. Three *narrative structure* tactics were also identified: *interactivity,* e.g. pop-ups, *ordering,* i.e. the ways of arranging the path viewers take through the visualisation, which can either be linear, random or user-directed,[3] and *messaging*. Messaging can include observations and commentaries to the viewer through short text fields such as labels, captions, headlines and annotations to add context, or more substantial descriptions such as articles, introductions, and summaries.

In our reading of the research into infographics, perhaps the most relevant work from an information architecture and cognitive slant is that of Ciuccarelli (2012), which calls for a stronger integration of narrative elements into visualisations in order to increase understanding. For this purpose, he defines narrative visualisations as *visual macroscopes*. Drawing on de Rosnay's (1975) idea of the macroscope as a metaphor symbolising a global approach to problems, systems and phenomena, the term *visual macroscope(s)* describes narrative visualisations as tools that draw the bigger context of the phenomena and "thus allow people to see not the infinitely

[2] The authors describe a cinematic scene represented through a visualisation to demonstrate how technology can provide new media and new genres with ways to present information in a story-like construct.

[3] The authors use ordering types to identify seven distinct genres of visual narratives, including magazine style, annotated charts, partitioned posters, flow charts, comic strips, slide shows, films/videos, animations, and then put them on a continuum of author-driven and reader-driven approaches.

small or the infinitely distant, as *microscopes* and *telescopes* do, but the infinitely complex" (Ciuccarelli 2012, p. 80). As the *macroscope* can help us enlarge our vision of the world, to better transmit knowledge, to free new values and act mindfully (de Rosnay 1975), similarly *visual macroscopes* offer opportunities to build a shared body of information, together with the possibility of sharing hypotheses and ideas (Ciuccarelli 2012); in this capacity, it is here argued that they represent discursive spaces for popularisation. Narrative visualisations use elements that go beyond the mere visualisation of data; thus, they work perfectly with socio-economic issues, "which are by nature complex, cannot be measured and can only be described in qualitative terms" (Ciuccarelli 2012, p. 79).

3. History of *infographics*

From a historical perspective, as mentioned above, infographics are not 'technically' new. As recounted by Rendgen (2012), history demonstrates that there is an ancient tradition of representing, producing and transmitting information, primarily scientific and technical, by means of visuals. Since the Middle Ages, in particular, visual representations have been used as teaching aids or important devices for conveying complex information. Examples include the cartographic tables and anatomical atlases of the Middle Ages, the cosmological theories and music annotations of the Renaissance, 17th century geography, and the statistical graphics used by 18th and 19th century officials and politicians, initially accessible only to small circles. In the 19th century, statistical patterns were also applied to historical subjects; examples include Florence Nightingale's diagram to display the mortality rate in British hospitals during the Crimean War and Minard's flow maps to visualise casualties and the narrowing flow of soldiers during Napoleon's catastrophic Russian campaign.

While these are good examples of pure visualisations and data visualisations, they are not all infographics in the current interpretation of the term, according to which illustrations, images and data visualisations – i.e. statistical graphics, tables and diagrams – are meant as a part of the elements used within the design of infographics (Rendgen 2012; Krum 2014). Notwithstanding these differences, *infographics* and (*data*) *visualisations* are often used interchangeably and to many, they continue to mean the same thing (Krum 2014).

The modern interpretation of the term *infographics* has evolved from its original application, according to which *infographics* were "ways of obtaining diagrams and maps from a computer (infographics)" (de Grolier 1979, p. 38)[4] as succinct and effective types of information delivery in the internal organisation of public administration and governmental authorities. This definition became the common and accepted interpretation of the term: *infographic* was simply "any visual representation of data or information", generally through a computer. Charts, graphs, data visualisations, maps, diagrams, and tables were all considered infographics. For many years, infographics were mostly confined to newspaper designers; *infographic* became a behind-the-scene term used in the print production of art departments, especially in the USA.

This use has been changed by the Internet: since the turn of the new millennium, people are now using infographics every day as their primary source of news and information, and their importance in social and digital media is indisputable. By adding a visual element, photos, illustrations, and charts have always improved the readability of texts and stories in print publications, and that phenomenon is even stronger in public spaces, thanks to the many available online publishing tools and the increased professional nature of graphic design.

Figure 1 represents an infographic summarising the historical evolution of the modern concept of *infographics*.

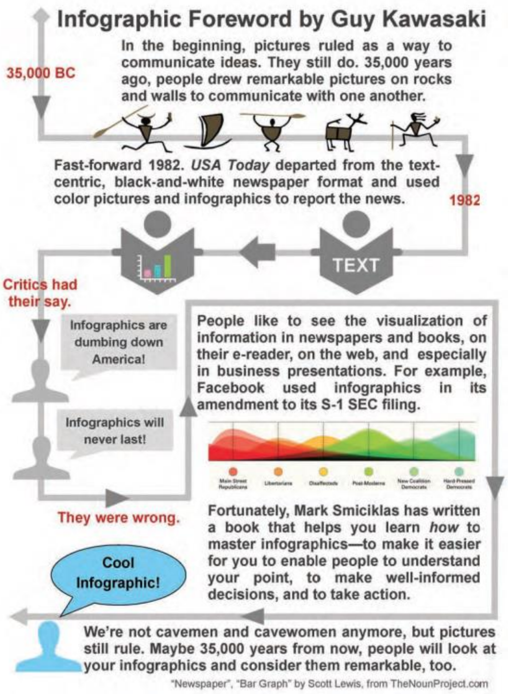


Figure 1
The historical evolution of the modern concept of infographic (Smiciklas 2012: xiii).

[4] "It matters little to decision-makers if it is a printed or stencilled document, a computer list, a diagram, map or photograph although, all things being equal, a decision-maker would prefer the most succinct" (de Grolier 1979, p. 38).

4. Infographics as a genre: layout, structure and communicative purposes

Segel and Heer (2010) and subsequent researchers in information visualisation have referred to narrative visualisations as genres, and have analysed them on the basis of varying strategies of visual narrative and layout, narrative structure and narrative rhetoric that emerge from visualisations across several media and different contexts. Other studies (e.g. Toth 2013) have analysed the pedagogical use and implications of infographics as a genre in professional communication.

According to the linguist Bhatia (1993, 1995), genres are instances of conventionalised or institutionalised textual artefacts in the context of specific institutional and disciplinary practices, procedures and cultures, and are identified on the basis of conventionalised features, typical textualisation patterns and socially recognised communicative purposes. Genre analysis has always been a multi-disciplinary activity, attracting attention not only from linguists, discourse analysts, communication experts and rhetoricians, but also scholars and professionals from other disciplines, all aiming to understand how members of specific discourse communities construct, interpret and use these genres to achieve their community goals. From an examination of the literature in various fields, we realise that genre analysis of visualisations and infographics has been terrain explored only by information scientists, communication scholars and professionals, artists and designers, and not by linguists, whether theoretical, applied or computational.

In the following sections we consider the three characteristics of infographics as a genre, namely visual layout, narrative structure and communicative functions from the perspective of graphic design. Krum's (2014) study is adopted, which, to our knowledge, represents the only systematic and comprehensive description available. As a data visualisation and infographics designer, however, Krum's study mainly reflects the concerns and practices in graphic design, and does not represent a linguistic genre analysis.

4.1. Layout

The first feature linked to genre is visual layout. In graphic design, the layout or format is referred to in various ways, such as a *tall, long* or *tower* and *horizontal* design (Krum 2014). According to Krum, the format chosen depends on one's publication strategy. From this perspective, the tall format (Figure 2) has some distinct advantages in online distribution, as most websites are designed to allow easy vertical but not horizontal scrolling; most infographics online follow this same format.

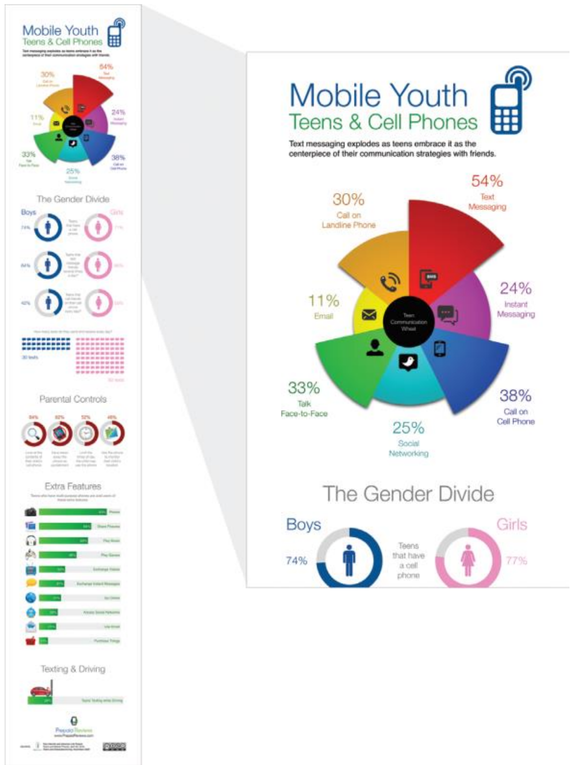


Figure 2
Tall format infographic (Krum 2014, p. 59).

In presentation slides and in printing, the tall format has some disadvantages, as it has to be scaled down to make its height fit the size of the page and this creates images that are usually too small to read or understand.
By contrast, the horizontal format (Figure 3) requires viewers either to scroll side to side to see all the content or to reduce the design to a very small size. As a result, tall format designs are much easier for the reader to view.

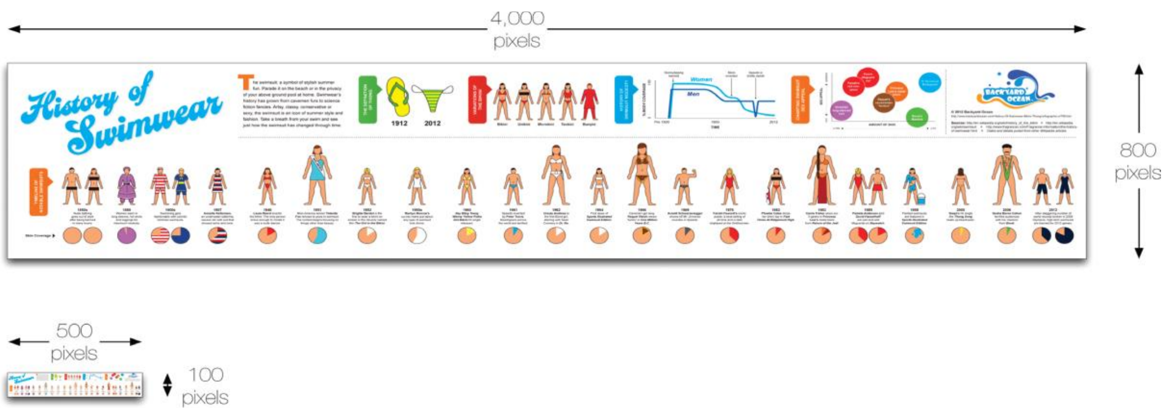


Figure 3
Horizontal format infographic (Krum 2014, pp. 62-63).

A very important aspect influencing choice of format in graphic design is information processing, which is not discussed by Krum.

4.2. Structure

The second feature linked to infographics as a genre is structure. In their capacity as narrative visualisations, many infographics have a narrative structure which follows a three-part sequential order, namely introduction, key message, and conclusion, each part having a proper function: appeal to the reader, comprehension and retention of information, respectively (Krum 2014).

The introduction presents the topic of the infographic, which is usually identified through some combination of the title and a brief paragraph of text, which has to be reduced to a minimum. This section also lays the foundation for the information before the reader moves to the key message. The key message represents the main event which is usually the dominant visual portion of the infographic through a large illustration or data visualisation.

Infographics also end with some closure, where the message is wrapped up for the reader, like the ending of a good speech or article. If appropriate, this is where a call-to-action should be included, if there is some type of follow-up that the readers are invited to take after reading the key message.

4.3. Communicative purposes

The third feature linked to infographics as a genre is communicative purpose. According to Krum (2014), the purpose of infographics can be assimilated to the same objectives as public speaking or writing: to inform, entertain, or persuade. In this sense, the taxonomy he proposes includes five distinct categories: *informative* infographics, *persuasive* infographics, infographic *advertisements, PR* infographics and *visual explanations*. The five categories will be briefly described.

4.3.1. Informative Infographics

Informative infographics are the dominant type of design, particularly online. After publishing infographics online for a number of years, Krum is convinced that if the goal of the design project is to maximise the number of views, visitors, and backlinks to the hosting site, informative infographics are more successful than all other types of designs. His theoretical stance is that online audiences are better disposed to read valuable information rather than advertisements.

An example of what Krum identifies as a good informative infographic is in Figure 4: the design has been published by an online backup company and focuses on the lifespan of storage media. Krum (2014, p. 71) observes that even though the company wants to sell its services to new customers, “the infographic isn’t a sales pitch”: the infographic uses data visualisations to convey the overall message that all storage devices are at risk of failure over time, but an online storage solution offers a backup that has some additional benefits. Then he concludes that “the message is meant to be independent and objective, [and that] the objective of informative infographics is to tie the value of the information presented to the value of the brand [which] creates a positive perception of the brand”, that is self-promotion. Krum does not support his claims with any linguistic insight.

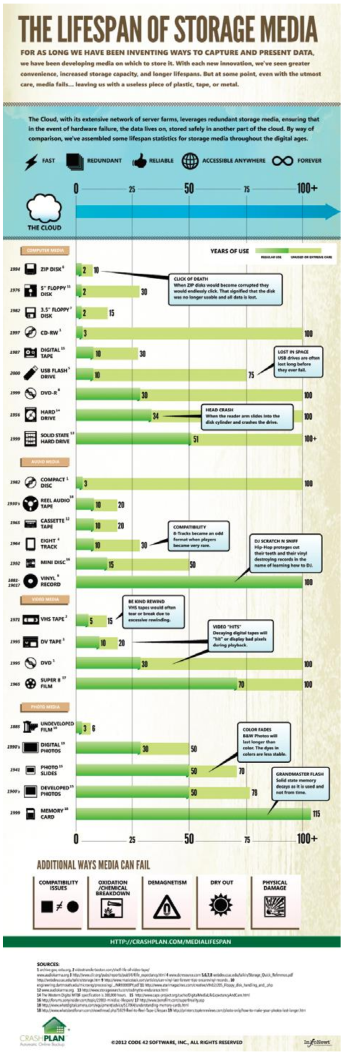


Figure 4
Informative Infographic (Krum 2014, p. 72).

4.3.2. Persuasive infographics and infographic advertisements

Persuasive designs are intended as a clear call to action and they attempt to convince readers to do something after seeing the infographic, such as performing good deeds or taking actions to help themselves, the local community, or the world at large. The actions might be to vote for a candidate/party/referendum, raise awareness of a health-related or environmental issue, visit a website, donate to a charitable organization, join a community, sign a petition, buy a product, to mention just a few.

Behind successful persuasive infographics is a narrative structure that leads the reader through the sequence of information: the technique Krum identifies is Monroe's *Motivated Sequence*, which motivates an audience to respond affirmatively to the speaker's purpose. The sequence contains five distinct steps. First, *Attention*: this step consists in getting the attention of the audience using a detailed story, shocking example, dramatic statistic or quotation, etc. Second, *Need*: this step describes the problem and explains how it applies to the psychological need(s) of the audience using statistics, examples, etc. and thus demonstrates a need for change. The premise here is that the audience's needs are what motivates action. Third, *Satisfaction*: the audience needs to solve the issue, so specific and viable solutions are offered that individuals or communities can implement to solve the problem. Fourth, *Visualisation*: this step tells the audience what will happen if the solution is implemented or does not take place. Monroe recommends being visual and detailed. Fifth, *Action*: the step tells the audience what action they can take personally to solve the problem.

Figure 5 shows an example of what Krum identifies as a good persuasive infographic from this perspective.

Persuasion is most visible in the form of advertising: a subcategory of persuasive infographics is therefore represented by infographic advertisements, which attempt to motivate the audience to take action such as purchase specific products or services. Persuasive infographics and infographic advertisements share the same communicative purpose: from a linguistic and socio-cognitive perspective, they belong to the same colony of promotional genre; what is not clear, therefore, is why Krum puts advertising infographics into a separate category.

Figure 5
Persuasive infographic (Krum 2014, p. 79).

4.3.3. PR Infographics

These are infographic designs particularly used by companies for public relations with press releases to build awareness of products and brands, provide information to shareholders, or increase the value of the brand.

4.4.4. Visual explanations

The final category is represented by visual explanations, which try to explain an idea or a process, such as how laws are made, relationships, or a complex concept.

5. Analysis and discussion

In the following sections we focus on the socio-semiotic, linguistic and discursive strategies adopted in a corpus of infographics from the World Health Organisation. The approach to infographics adopted for this analysis is Ciuccarelli's (2012) interpretation of narrative visualisations, hence infographics, as *visual macroscopes*.[5]

Two layers of analysis are adopted: firstly the social semiotic, i.e. the use of different modes and their interrelations in their capacity to create all three of Halliday's meaning functions (Kress, van Leeuwen 2001; Kress 2010); secondly, the linguistic and discursive, i.e. the interrelations between the cognitive dimensions arising from the corpus data and the linguistic and discursive strategies that realise them (Moirand 2003). The linguistic and discursive structures that specifically interest us include the set of metalinguistic expressions that introduce specialised terms and information, along with instances of explanation through paraphrase and exemplification.

5.1 The corpus

The data for the present analysis are taken from a corpus of six infographics produced by the Department of Nutrition for Health and Development of the World Health Organisation[6] and addressed to WHO Member States. These infographics are part of a comprehensive implementation plan which aims to achieve six global nutrition targets in maternal, infant and young child nutrition through direct nutrition interventions and multi-sectorial actions in the food system, education and social protection by 2025. These global nutrition targets include: reducing anaemia and improving breastfeeding in maternal nutrition, as well as reducing low birth weight, stunting, wasting and excessive weight gain in infant and young child nutrition.

5.2. Socio-semiotic analysis

This section analyses the ways in which meaning is socially and culturally constructed and interwoven in the multimodal discourse of WHO infographics through various semiotic resources. One of the key premises of multimodal social semiotics is in fact that meaning-makers, including authors and graphic designers, always draw on a multiplicity of modes – i.e. a set of socially shaped and culturally given resources for meaning-making (Kress, van Leeuwen, 2001; Kress 2010) – to make meanings. These modes are arranged in a multimodal

[5] See Section 2.

[6] The infographics are available at https://www.who.int/nutrition/global-target-2025/infographics/en/.

design. Selections are made in each mode and each mode offers ways to highlight that to which the audience's attention is to be drawn. This is how text makers can create reading paths and shape how the audience navigates the text and process information. In this sense, mode is "an organizing principle of presentation and communication and therefore treated as a central unit of analysis" (Bezemer, Jewitt 2010, p. 183).

The WHO uses layout and pictorials, but also typography – i.e. choice and size of fonts, margins, hierarchy between headings, subtitles and background text –, data visualisations and writing as modes of representation: they regularly combine and represent well-acknowledged regularities through which the storytelling is carried forward, and thus help the audiences locate the information and identify the important information.

In the following subsections, we will analyse how these resources are used to fulfil a variety of text functions.

5.2.1. Layout and structure

All the WHO infographics in the present corpus display the same vertical format and layout and the same structure: information processing and reading path progress from top left to bottom right, according to the Z-reading pattern typical of the western world (Kress, van Leeuwen 1996). This pattern is the most immediate and natural and helps define the *zones* where the different visual elements, and corresponding information, are set and thus identify the narrative pattern. The infographic on *stunting* (Figure 6) is used as a paradigm.

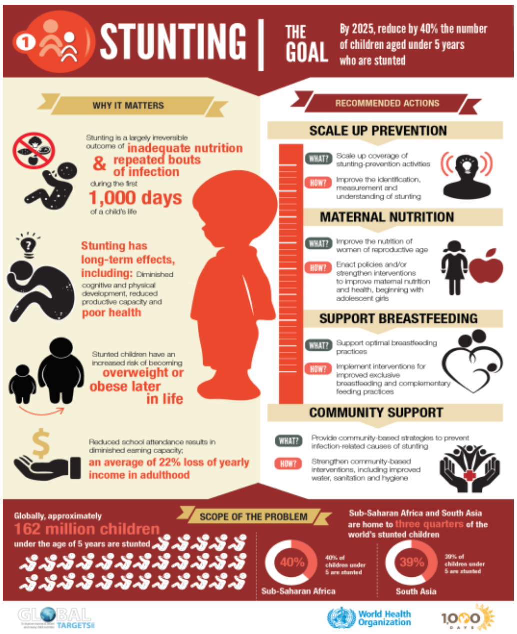


Figure 6
Stunting.

The elements of all WHO infographics are located according to an *Ideal* (top)/*Real* (bottom), *Given* (left)*/New (right)*, *Centre/Margins* triptych, even though all the infographics do not conform to a traditional Z-shape layout, as the number of Z-angles along the reading path varies. The information positioned on the top is *The Goal* and represents the *Ideal* part, i.e. representation of what the world should be like (Kress, van Leeuwen 1996): in this corpus, it corresponds to reducing anaemia in maternal nutrition, as well as reducing low birth weight, stunting, wasting and excessive weight gain in infant and young child nutrition. The white headline draws the attention first against the dark background. From the left, where the topic is given, the path leads across to the new information on the right and this is the first horizontal line. The top horizontal line includes the main components the Organisation wants viewers to focus on first.

From here the path continues diagonally over to the left of the following part. This part occupies the *Centre* and features any information that builds up the discourse: this is the first Z-angle. From the *Why it matters* section on the left, the path continues horizontally over the right side to the *Recommended actions*, and this is the second horizontal line. Following the path, attention is then directed to the left of bottom part: this section is the *Scope of the Problem* and represents the *Real,* i.e. shows what the world is really like. This is the second Z-angle. From here, the reading path moves horizontally to the right and this represents the second and final horizontal line.

The two central sections are visually organised through subsequent sub-sections, separated and introduced by pictorials and text in the left section – i.e. *Why it matters* – and by headings in the section on the right – i.e. *Recommended actions* –: this section displays an arrangement dominated by writing and its structure resembles the textual organisation of a book in chapters – represented here by sub-sections –, and paragraphs – here represented by "What?" and "How?" sub-sub-sections –, with pictorials playing a marginal role. Pictorials and text thus function as text organisers, but occupy different roles across the two sections, as will be explained in the following sections.

The narrative pattern develops along the reading path and moves from the Ideal part at the top through the two sections in the centre to the Real part at the bottom. This pattern appears as a slightly modified version of Monroe's *Motivated Sequence* in persuasive infographics, with the step *Action* preceding *Attention* and the step *Visualisation* implicitly implied from *Attention*.

In the next sections we will focus on ways in which colour, typography, pictorials and writing or text support the narrative structure and realise the persuasive function of the WHO infographics.

5.2.2. Colour, typography, images and data visualisations

In this section we analyse how colour, typography, pictorials, data visualisations are used to articulate the content matter and help the reader locate the important information.

For the purposes of this section, the infographics on stunting (Figure 6) and anaemia (Figure 7) have been chosen as a paradigm.

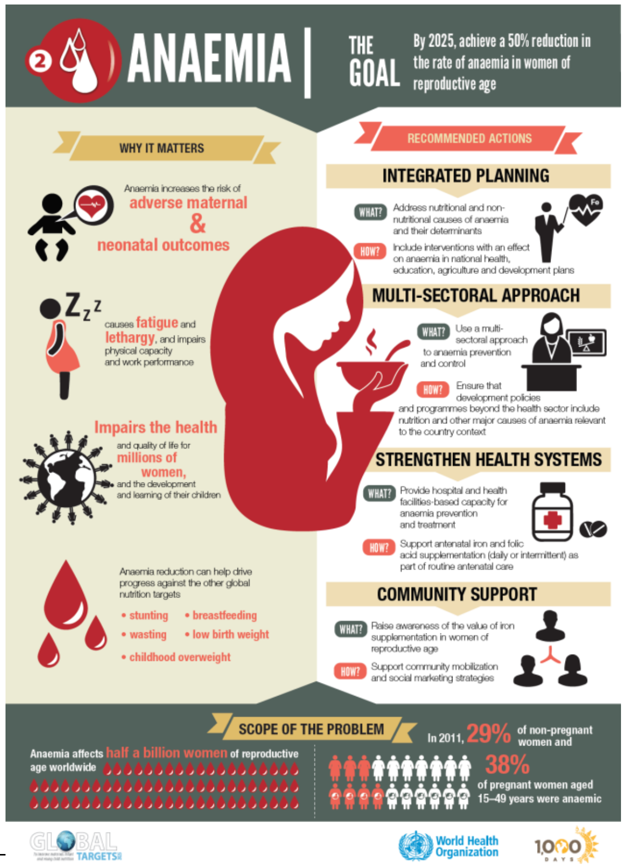


Figure 7
Infographic on anaemia.

The different modes often combine and contribute to the information organisation and narrative structure in different ways and according to different patterns: while modes such as colour, choice and size of font are exploited throughout the whole text, others such as pictorials, data visualisations, numbers headings and subtitles are predominant only in specific parts.

In *The Goal* part, the headline is highlighted through the use of white against the dark background; font size is also used to highlight the words referencing the topic of the whole design on the left and its *Goal* on the right. The section is therefore a call to action.

In the *Why it matters* section, colour, choice and size of font draw attention to the words that indicate the different topic-oriented discourses and how they are linked to each other, like the physical outcomes of stunting (Figure 6) in infants (e.g. "repeated bouts of infection, 1,000 days") and children ("long-term effects, poor health"), but also in adult life (e.g. "a higher risk of becoming overweight or obese later in life"), in addition to its social and economic outcomes in adulthood (e.g. "an average of 22% loss of yearly income in adulthood"). Similarly, in the same section of *Anaemia* (Figure 7), these modes draw attention to the words referencing the physical outcomes of anaemia, both neonatal and maternal (e.g. "maternal & neonatal outcomes; fatigue, lethargy"), in addition to the size of the problem (e.g. "impairs the health [of] millions of women") and how its reduction helps drive progress against other global nutrition targets (e.g. "stunting, breastfeeding, wasting, low-birth weight and childhood overweight"). In addition, pictorials introduce and visualise the different topic-related discourses − medical, social and economic, to mention a few: the selection and interpretation of their meanings, both literal and symbolic, is directed by the following verbal text. The relation between pictorials and text is thus of an anchorage type (Barthes 1977), as they both trigger the bigger context behind the topic and its infinite complexity. Each discourse within this context, then, represents how the topic applies to the psychological needs of the audiences, thus offering motivations for change.

In the right section on *Recommended actions*, choice and size of font draw attention to the headings referencing viable solutions to the problem, in response to the motivations highlighted in the previous section: the headings include "Scale-up prevention", "Community Support", and Support Breastfeeding (Figure 6); "Integrated planning", "Multi-sectorial approach", "Strengthen health-systems", and "Community support" (Figure 7). Each section is further subdivided in subsections introduced by the colour-coded headings "What?" and "How?", which detail what the viable solutions are and how they can be implemented. Compared to the left section, here the pictorials play a marginal role and mainly summarise the topic. In this section, pictorials, visual graphic, typography and text combine to simplify the complexity of the topic.

In the *Scope of the problem*, colour, choice and size of font combine with text and data visualisations like numbers, percentages and icons to indicate the extent of the problem either globally or in specific areas: here, the text provides a description of the process and the entities involved in the problem, while numbers, percentages and icons visualise them. Without the use of either the one or the other, the information provided by the written account or the visualisations alone would be severely limited. In addition, since this part presents the current state of the problem, there is an interest in grounding the information scientifically in a quantitative approach, bringing 'scientific evidence' from various disciplines and with 'scientific discourse' as a framing.

It can be concluded that layout, visual graphic, typography and writing are semiotic resources which realise all three of Halliday's meaning functions, which is one of the premises of multimodal social semiotics. In the WHO infographics analysed, layout, structure – i.e. sections and sub-sections –, writing, headings, subtitles and pictorials combine to articulate content matter; colour, choice and size of fonts and data visualisations in the form of numbers, percentages and icons highlight the salience of each piece of information; all together, these elements help to lead the interpretation of the text from one element to another in order of decreasing impact; in short, they assist in concept development. Pictorials, in particular, are used in several of the following capacities, generally acknowledged in the visual communication of science (Trumbo 1999, 2000): they illustrate concepts, assist in concept development and summarise a topic.

5.3. Linguistic and discursive analysis

In the following sections, we analyse how the linguistic and discursive strategies support the narrative structure of the WHO infographics, as well as how they assist in concept development and trigger the Visual Macroscope: the approach adopted is bottom-up, from the lexical to the grammatical, textual, and then discursive level. In this section, examples from infographics other than stunting and anaemia are also included.

The "Goal" section (e.g. in Figure 7 "By 2025, achieve a 50% reduction in the rate of anaemia in women of reproductive age") deals with the lexico-semantic field of reduction (e.g. "achieve a […] reduction in the rate of […]" or "reduce (by + percentage) the number of") and is realised by non-finite verb phrases or elliptical noun phrases, as in "no increase in childhood overweight", seen in the infographic on overweight children All these clauses are marked and start with the time adjunct "by 2025", which represents the textual theme. The local lexico-grammar, therefore, indicates that the section is a call-to-action.

The lexico-grammatical patterns in *Why it matters* (e.g. "Anaemia increases the risk of adverse maternal and neonatal outcomes", Figure 7; "Stunting is a largely irreversible outcome of inadequate nutrition and repeated bouts of infection during the first 1,000 days of a child's life", Figure 6) correspond to unmarked declaratives, where the nutrition issue discussed is the subject occupying the thematic position, and the rhemes consist in finite verb phrases with modification, telling how the problem applies to the psychological needs of the audience(s), according to the following patterns: X (i.e. the problem) + VP + NP [+PP]/PP. Sometimes elliptical subjectless declaratives also occur and exemplify instantiations of textual ellipsis as a cohesive device (e.g. “causes fatigue and lethargy, and impairs physical capacity and work performance”, in the infographic on anaemia, Figure 7): the omitted subject is the problem discussed and can be recovered from the previous co-text. All these patterns exemplify that information structure is mainly realised through thematic progression with a constant theme (e.g. “Stunting is a largely irreversible outcome of inadequate nutrition and repeated bouts of infection during the first 1,000 days of a child’s life”; “Stunting has long-term effects including: […]”, Figure 6; “Anaemia increases the risk of adverse maternal and neonatal outcomes”, “[Anaemia] causes fatigue and lethargy, and impairs physical capacity and work performance”, “[Anaemia] impairs the health and quality of life of millions of women […]”, Figure 7). This foregrounds the fact and idea that the problem discussed is a priority to the WHO Member States. These lexico-grammatical patterns also realise the discursive strategy of explanation, which is the most frequently used strategy in popularisation discourse (Moirand 2003).

How the problem applies to the needs of the audience(s) is explained through different topic-oriented discourses, which correspond to the fundamental underlying categories that structure the properties of the issue discussed, such as Cause (e.g. “stunting is a largely irreversible outcome of inadequate nutrition and repeated bouts of infection”, Figure 6), Consequence (e.g. “stunting has long-term effects”, Figure 6; “childhood overweight increases the risk of obesity […]”, seen in the infographic on overweight children; “low birth-weight is a major predictor of perinatal mortality”, in the infographic on low birth weight; etc.), Quantity (e.g. “anaemia impairs the health and quality of life of millions of women […]”, Figure 7) and Localisation, either geographic (e.g. “low-birth weight is a global concern”, in the infographic on low birth weight*;* “childhood overweight is increasing in all regions of the world”, in the infographic on overweight) or socio-demographic (e.g. “Majority of low birth weight births occur in low-income countries”, in the infographic on low birth weight).

In addition, the causal relation, which includes both cause and consequence, is more extensively represented and structured from a discursive perspective than the other two categories: its lexical units reveal that, while the cause is generally identified with inadequate nutrition or another disease/physical problem, the consequences can be multifaceted: physical (e.g. “children who are overweight or obese are at a higher risk of developing serious health problems; wasting increases risk of stunted growth […]", in the infographic on overweight; "low birth weight increases the risk for non-communicable diseases such as diabetes and heart disease", in the infographic on low birth weight, etc.), cognitive (e.g. "stunting has long-term effects including: […] diminished cognitive development", Figure 6; "breastfeeding provides babies everything they need […] for brain development", infographic on breastfeeding), and social, particularly at the level of work performance and earning capacity in adulthood ("stunting has long-term effects, including: […] reduced productive capacity"; "anaemia impairs […] and work performance, Figure 7). In addition, when the health-related consequence is represented either through a general term, as in "Stunting has long-term effects" (Figure 6), or a specialised one, as in "Long-term effects including: […] non-communicable diseases" (Figure 6), exemplification is used, and it is introduced by metalinguistic expressions like *such as* or by means of punctuation, normally a colon (:).

These global targets are also linked to each other and interdependent: a reduction in anaemia, for instance, "can help drive progress against the other global nutrition targets: stunting, breastfeeding, wasting, low birth weight, childhood overweight" (Figure 7).

The overarching discourses – medical, scientific, social and economic and a mixture of these – regulate the infographic design and reflect the interests and concerns of the WHO and its epistemological positions. In the examples discussed, the discourses are realised through repetitions of some contextual lexico-grammatical patterns, especially in the discussion of the health-related consequences. Examples include the repetition of the verb pattern "X – i.e. the problem or person affected by it – has a high risk of developing […]" or the noun phrase "non-communicable diseases […]" as in the infographics on stunting and overweight. The design foregrounds an interest in the multifaceted nature of the problems discussed (primarily nutrition and medical issues) which, in addition to being closely related to each other, can also cause other health-related and cognitive problems, in turn likely to develop into social and economic issues later in life.

These global targets also need a multifaceted approach, which is organised into several sections within *Recommended actions* on the right side of the infographic. The lexical patterns of these section headings reference a variety of approaches: health, education and development planning (e.g. "Integrated Planning", Figure 7), disease prevention (e.g. "Scale-up Prevention", Figure 6), health systems (e.g. "Strengthen health systems", Figure 7), research (e.g. "Build up evidence"), social protection (e.g. "Community Support", Figure 6 and 7), legal and economic protection (e.g. "Support paid leave", in breastfeeding), education ("Dietary guidelines", in overweight), marketing of food substitutes (e.g. "Limit formula marketing of breastfeeding substitutes”), etc. Within each section are two sub-sections, “What?” and “How?”, which explain the viable solutions and how they can be implemented. They also showcase the interrelationship among overarching discourses – medical, legal, policy, social, education, scientific research, etc.: e.g. “Include interventions with an effect on anaemia in national health, education, agriculture and development plans to address causes of diseases” (Figure 7); “Enact policies to enhance food systems to support dietary practices” and “Develop public and social marketing campaigns to support regulations of food marketing” in the infographic on overweight; “Rapidly develop evidence to reduce the burden of wasting, which can then be translated into policy actions”, in the infographic on wasting.

The lexico-grammatical patterns in the Recommended actions include non-finite VPs and NPs with complex modification and indicate that thematic progression is realised through subsequent themes derived from the hyper-theme. The hypertheme is the heading; “What?” and “How?” are the subsequent themes and the rhemes are the non-finite VPs or NPs that follow. The examples discussed indicate that the explanations in this section involve complex discourses, which are realised through different subsequent themes, with higher lexical diversity and density than any other section of the design.

Some considerations emerge from our analysis of WHO infographics. From a socio-semiotic perspective, the different modes used in the corpus correspond to those exploited in many other multimodal genres. At a linguistic and discursive level, however, the lexico-grammatical patterns trigger a variety of overarching and interrelated discourses or interdiscourses, which exemplify the infinite complexity of socio-economic issues such as health and nutrition. Their interrelationship turns these infographics into “visual macroscopes” (Ciuccarelli 2012) of the complex relation between science, nature and society. Within the WHO's implementation plan for the six global target issues – i.e. stunting, anaemia, low birth weight, overweight, breastfeeding and wasting –, these interdiscourses serve a variety of functions: they occur as “reminders” (Moirand 2003) of prior discourse moments within the global target issues by creating an *interdiscursive memory bank* (Moirand 2003) which enables the audiences to move backwards and forwards within these issues; they also act as constant “reminders” of the meaning-maker’s – i.e. the WHO – position and interests.

6. Conclusions

Infographics are pliable discursive spaces which serve the process through which scientific and specialised knowledge is spread across a variety of contexts (e.g. medicine and health, business, environment, etc.) and socio-linguistic levels of communication (e.g. among experts in science magazines, between expert professionals and lay audiences in journalism and business, for example). From the survey of the literature on infographics, we conclude that as a genre of popularisation it invites interest also from a socio-semiotic and linguistic point of view.

The examples discussed in the purposely constructed corpus illustrate how infographics are used in the context of important and global socio-economic issues to facilitate information transfer to government bodies, which can then implement informed policy actions: infographics gather specialised information produced by different discourse communities, and integrate their discourses –i.e. medical, legal, social, educative, etc. – into a complete and complex whole. The integration of narrative visualisation (the concept of "visual macroscopes") and semiotic resources into an infographic enables the interrelationship between discourses to be seen and made sense of. It is our view that infographics thus have considerable educational potential as a genre, and, given the visual age we live in, could be exploited more in different fields, including, for example, that of linguistics.

The present experimental analysis was deliberately conducted on a small corpus, which revealed specific socio-semiotic, lexico-grammatical and discursive patterns. While valid in their own terms, the results invite further investigation in larger and more diversified corpora.

Bionote: Sonia Piotti is Research Fellow of English Language at the Department of Modern Languages, Università Cattolica in Brescia. She holds a PhD in "Linguistica Applicata e Linguaggi della Comunicazione" from Università Cattolica. Her research interests lie in the fields of the History of the English Language, English Lexicology and popularisation discourse in English. She is currently doing research into the medical vocabulary of Old English and the popularisation of mathematical sciences in Early Modern English.
Amanda Murphy (BA Hons Cantab; Laurea quadriennale, Università degli Studi di Milano; PhD Birmingham, UK) comes from a background of Modern Languages (French and Italian) and Applied English and Italian Studies. As a Professor of English and translation at Università Cattolica del Sacro Cuore, she conducts research on corpus-assisted analyses of specialized language, English as a medium of instruction in international education, and language and art associated with the Roman Veronica.

Authors' addresses: sonia.piotti@unicatt.it; amanda.murphy@unicatt.it

References

Barthes R. 1977, *Image, Music and Text. Roland* Barthes, Essays selected and translated by Stephen Heath, Fontana Press, London.

Bhatia V.K. 1993, *Analysing Genre: Language Use in Professional Settings*, Longman, London.

Bhatia V.K. 1995, *Genre-mixing in Professional Communication: The Case of 'Private Intentions' v. 'Socially Recognised Purposes'*, in Bruthiaux P., Boswood T. and Bertha B. (eds.), *Explorations in English for Professional Communication*, City University of Hong Kong, Hong Kong, pp. 1-19.

Bezemer J. and Jewitt C. 2010, *Multimodal Analysis: Key Issues*, in Litosseliti L. (ed.), *Research Methods in Linguistics*, Continuum, London, pp. 180-197.

Calsamiglia H. and Van Dijk T.A. 2004, *Popularization Discourse and Knowledge about the Genome*, in "Discourse and Society" 15 [4], pp. 369-389.

Ciuccarelli P. 2012, *Turning Visualisations into Stories and 'Big Pictures'*, in Rendgen S., *Information Graphics*, with Essays by Paolo Ciuccarelli, Richard Saul Wurman, Simon Rogers, Taschen, Köln, pp. 79-84.

Ghadessy M., Henry A. and Roseberry R.L. 2001, *Introduction*, in Ghadessy M., Henry A. and Roseberry R.L (eds.), *Small Corpus Study and ELT. Theory and Practice*, Benjamins, Amsterdam/Philadelphia, pp. xvii-xxii.

Gershon N. and Page W. 2001, *What Storytelling Can Do for Information Visualization*, in "Communications of the ACM" 44 [8], pp. 31-37.

De Grolier É. 1979, *The Organization of Information Systems for Government and Public Administration*, Unesco, Paris.

The Institute for the Advancement of Research in Education (IARE) at AEL 2013, *Graphic Organizers: a Review of Scientifically Based Research*. http://www.inspiration.com/sites/default/files/documents/Detailed-Summary.pdf (13.03.2018)

Kress G. and van Leeuwen T. 1996 (2006), *Reading Images: The Grammar of Visual Design*, Routledge, London.

Kress G. and van Leeuwen T. 2001, *Multimodal Discourse: The Modes and Media of Contemporary Communication*, Arnold, London.

Kress G. 2010, *Multimodality. A Social Semiotic Approach to Contemporary Communication*, Routledge, London.

Krum R. 2014, *Cool Infographics: Effective Communication with Data Visualization and Design*, John Wiley & Sons, Indianapolis.

Lankow J., Ritchie, J., and Crooks R. 2012, *Infographics: The Power of Visual Storytelling*, John Wiley & Sons, Hoboken.

Miller G. 1983, *Informavores*, in Machlup F. and Mansfield U. (eds.), *The Study of Information: Interdisciplinary Messages*, John Wiley, New York, pp. 111-113.

Moirand S. 2003, *Communication and Cognitive Dimensions of Discourse on Science in the French Mass Media*, in "Discourse Studies" 5 [2], pp. 175-206.

Oxford English Dictionary. http://www.oed.com (3.4.2018).

Rendgen S. 2012, *Introduction*, in Rendgen S., *Information Graphics*,with Essays by Paolo Ciuccarelli, Richard Saul Wurman, Simon Rogers, Taschen, Köln, pp. 9-18.

de Rosnay J. 1975, *Le Macroscope. Vers une Vision Globale*, Seuil, Paris.

Segel E. and Heer J. 2010, *Narrative Visualization: Telling Stories with Data*, in "IEEE Transactions on Visualization and Computer Graphics" 16 [6], pp. 1139-1148.

Sinclair J.M. 2001, *Preface*, in Ghadessy M., Henry A. and Roseberry R.L. (eds.), *Small Corpus Study and ELT. Theory and Practice,* Benjamins, Amsterdam/Philadelphia, pp. vii-xv.
Smiciklas J.M. 2012, *The Power of Infographics. Using Pictures to Communicate and Connect with Your Audiences*, QUE Pearson Education, Indianapolis.
Toth C. 2013, *Revisiting a Genre: Teaching Infographics in Business and Professional Communication Courses*, in "Business Communication Quarterly" 76 [4], pp. 446-457.
Trumbo J. 1999, *Visual Literacy and Science Communication*, in "Science Communication" 20 [4], pp. 409-425.
Trumbo J. 2000, *Seeing Science. Research Opportunities in the Visual Communication of Science*, in "Science Communication" 21 [4], pp. 379-391.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 315-336
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p315
http://siba-ese.unisalento.it, © 2019 Università del Salento

LES SLOGANS DE L'ÉLECTION PRÉSIDENTIELLE FRANÇAISE DE 2017: DES "PETITES PHRASES" POTENTIELLES?[1]

ALIDA MARIA SILLETTI
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Abstract – This research deals with political slogans widespread during the French presidential elections campaign in 2017 and with the way they circulate in the web. Based on an approach inspired to studies about enunciation and French political discourse analysis, we will identify political slogans presented by the eleven candidates to French presidential election and put their full words in the web to see how journalists and political opponents use these slogans to talk about the candidate who carries them. Our attention will be restricted to slogans that seem to be more suitable for an alteration of their original form and meaning in terms of simple formulation to create a joke or rather of a real modification of their message in a polemic attitude and modifying the candidate's message. Hence, we will analyze these alterations and examine them in order to see if they show the characteristics of a potential "petite phrase", that is a phenomenon typical of politics which originates from real declarations pronounced by politicians or by their staff and then widespread by media in a way of changing their original meaning. In particular, our analysis will concern political slogans by François Asselineau, François Fillon, Marine Le Pen and Emmanuel Macron and their alterations by blogs, press articles and speeches of their opponents. As a result, we will show that some deformations receive a greater attention and that each alteration has to be considered in its context, which is fundamental to understand the reasons and the way an original slogan is deformed.

Keywords: political slogans; presidential elections; "petite phrase"; meaning alterations; media

1. Introduction

Le discours politique vise à "s'adresser au plus grand nombre pour le faire adhérer à des valeurs communes" (Charaudeau 2005, p. 187) et cette intention se traduit par une médiatisation des thèmes abordés par les sujets politiques qui s'en emparent. Parmi les outils de simplification et de

[1] Nous remercions nos deux relecteurs anonymes pour leurs remarques et pour leur relecture critique de cet article.

raccourcissement de la parole politique, cette recherche portera sur le slogan politique en termes de conception et réalisation, de présentation au public, de lien avec le candidat qui le lance comme signe distinctif de son idéologie (Reboul 1975; Barbet 2012; Nianxi 2008). A partir de l'approche théorique de l'analyse du discours politique français (Charaudeau 2005), notre attention sera focalisée sur les slogans de la campagne pour l'élection présidentielle en France de 2017. Notre corpus, composé de 70575 mots, comprend tant les slogans des candidats que les commentaires sur les slogans diffusés dans le web et publiés par les journalistes, par des adversaires politiques, des politologues, mais surtout par les internautes. La manière dont les slogans sont appréhendés et médiatisés sur la Toile nous permettra de percevoir leurs détournements, parmi lesquels nous chercherons à identifier de potentielles "petites phrases" plus ou moins soudées aux candidats qui les portent. Une "petite phrase" est un syntagme dénominatif non savant relevant d'un "discours autre approprié" (Authier-Revuz 1996, p. 112) et possédant une valeur dénominative, un sens locutionnel et une stabilité qui lui permet d'être lié à un référent spécifique (Krieg-Planke 2011, p. 25). Pour l'analyse des textes du corpus, nous nous servirons également des outils offerts par la linguistique de corpus afin d'isoler les slogans et leurs détournements, leurs attestations au fil du discours et la manière dont ils sont exploités par leurs utilisateurs.

Après avoir rappelé les caractéristiques du slogan de propagande politique, nous présenterons les seize slogans des candidats du premier et du second tour de la campagne des élections présidentielles de 2017. Nous travaillerons par la suite sur le corpus des commentaires sur ces slogans pour comprendre à quoi les slogans politiques originels et les slogans détournés peuvent être apparentés et pour vérifier si ces derniers sont susceptibles de devenir de potentielles "petites phrases" de la politique française. Pour ce faire, nous isolerons sept slogans, que nous examinerons au sein des sous-corpus qu'ils engendrent.

2. Le slogan de propagande politique

Un slogan est "une formule concise et frappante, utilisée [entre autres] […] par la propagande politique".[2] En termes généraux, un slogan est bref et syntaxiquement simple (Reboul 1975; Jaubert 1985); il est caractérisé par des sonorités et des répétitions visant à faciliter sa mémorisation, et son interprétation recèle une portion d'implicite. Il est proche d'une devise en tant

[2] *Petit Robert 2017*, version numérique.

qu'expression d'une pensée, d'un sentiment, d'un mot d'ordre, mais il ne s'identifie pas avec cette dernière.[3]

Dans le domaine politique, un slogan relève de la sphère publique et vise à mobiliser des masses; il sous-tend une idéologie dont il cumule les éléments (Nianxi 2008), mais il fait également référence à une culture partagée par les publics visés pour qu'il soit non seulement compris et mémorisable par ceux-ci, mais aussi connu. Un slogan peut ainsi emprunter à la publicité commerciale ses méthodes et techniques (Barbet 2012) étant donné la visée d'influence du discours propagandiste. Tout comme ce dernier discours, un slogan recèle également une visée d'incitation par laquelle le sujet émetteur fait croire au destinataire être le bénéficiaire de ce qui lui sera proposé, tandis que le destinataire de l'énonciation doit croire être le bénéficiaire de l'action proposée. Un autre trait essentiel du discours propagandiste et du slogan politique est d'être toujours situé dans un "à-venir" (Charaudeau 2009, p. 25), de porter sur des événements présents mais dont les effets se percevront si certaines conditions sont remplies. Ces conditions relèvent de l'adhésion du public cible aux contenus du discours de propagande politique présenté dans le slogan et donc à celui qui s'en charge, à savoir le "porteur" du bien-être collectif annoncé par le projet futur. Le discours propagandiste est donc aussi un discours de promesse, prophétique et d'"incantation magique" (Charaudeau 2009, p. 23) car il concerne des descriptions évaluatives globales plutôt que des faits ou des connaissances. Nous ajoutons à ces remarques qu'un slogan a une nature événementielle puisqu'il est produit à un certain moment pour atteindre un but précis. Il peut rassembler les idées d'un parti ou mouvement politique depuis sa création et pendant sa "vie politique" ou une période de celle-ci, ou bien il peut être conçu et produit à une occasion particulière, telle qu'une campagne électorale.

En termes d'analyse du discours, les slogans sont des unités microlinguistiques car ils concernent d'abord le morphème et la phrase plutôt que le texte (Adam 2008).[4] Ils sont pourvus d'un co(n)texte plus ou moins élargi représenté aussi bien par leur matérialité langagière environnante – leur contexte de production – que par l'extralangagier – leur impact visuel en termes de format et taille des caractères, de couleurs, de combinaison entre le langagier et le graphique. Ces éléments, auxquels concourt également l'*ethos*[5] du sujet parlant, contribuent à rendre un slogan politique efficace (Maingueneau 2002). L'enjeu du discours politique consiste en effet à

[3] Selon leur étymologie, si "slogan" dérive de l'écossais, notamment du gaélique "cri (*gairm*) d'un clan (*sluagh*)", une devise se caractérise par le fait d'être un signe distinctif.

[4] Au contraire, une unité macrolinguistique est composée de séquences et de textes (Adam 2008).

[5] Suivant Amossy (2002, p. 238), nous définissons l'*ethos* comme l'"image de soi que le locuteur construit dans son discours pour exercer une influence sur son allocutaire".

affirmer l'identité d'un orateur – locuteur individuel ou collectif – qui se charge du rôle de "leader charismatique" (Charaudeau 2015) pour que son auditoire s'y identifie comme public de militants et/ou de potentiels électeurs (Mayaffre 2003).

Les slogans que nous étudions relèvent de la campagne pour l'élection présidentielle en France. Cette campagne est lancée depuis la circulation des candidats potentiels, elle se poursuit par leur candidature et elle se termine par les élections.

3. Les slogans politiques de la campagne présidentielle française de 2017

Notre corpus se compose de quatorze slogans accompagnant la campagne des candidats du premier tour de l'élection présidentielle française de 2017, et de deux slogans des candidats du second tour, parus dans l'entre-deux-tours. Tous les slogans montrent une identité dialogique entre texte et image car ils accompagnent le candidat lors de ses discours et meetings de campagne, ils sont prononcés par celui-ci, mais ils prennent également la forme d'affiches de chaque candidat à côté de son nom, de son visage et de son éventuel logo. Certains sont plus connus que d'autres du fait de la notoriété d'un candidat, de ses positions, dans un sens aussi bien positif que négatif: la circulation des slogans sur le web contribue à accroître leur visibilité et, à l'inverse, leur méconnaissance se traduit par leur moindre diffusion. Les médias sont donc un vecteur essentiel de médiatisation des slogans, ce qui contribue à leur diffusion et à leur mémorisation, mais également à leur possible détournement par les journalistes et les utilisateurs du web.

La plupart des slogans analysés relèvent de la seconde moitié de 2016. Certains sont plus récents: tel est le cas du second slogan de François Fillon, "Une volonté pour la France", paru en pleine affaire "Penelopegate" en février 2017, et des deux slogans de Marine Le Pen et d'Emmanuel Macron publiés dans l'entre-deux-tours, fin avril-début mai 2017.

(1) *Faire entendre le camp des travailleurs* (Nathalie Arthaud)[6]
(2) *Le candidat du Frexit* (François Asselineau)
(3) *Le parti qui monte malgré le silence des médias* (François Asselineau)
(4) *Libérez-nous de l'occupation financière !* (Jacques Cheminade)
(5) *Ni système, ni extrêmes* (Nicolas Dupont-Aignan)
(6) *Le courage de la vérité* (François Fillon)
(7) *Une volonté pour la France* (François Fillon)
(8) *Faire battre le cœur de la France* (Benoît Hamon)
(9) *Je veux retaper la France* (Jean Lassalle)

[6] Notre présentation des slogans suit l'ordre alphabétique des noms des candidats.

(10) *Au nom du peuple* (Marine Le Pen)
(11) *Choisir la France* (Marine Le Pen)
(12) *En Marche ! : La France doit être une chance pour tous* (Emmanuel Macron)
(13) *Ensemble, la France !* (Emmanuel Macron)
(14) *L’Avenir en Commun* (Jean-Luc Mélenchon)
(15) *La force du peuple* (Jean-Luc Mélenchon)
(16) *Nos vies, pas leurs profits !* (Philippe Poutou).

Du point de vue morphosyntaxique, ces slogans se composent d’un groupe nominal étendu, d’un groupe prépositionnel, d’un groupe adverbial ou bien d’une phrase s’ouvrant par un verbe à l’infinitif qui joue le rôle de nœud verbal de la phrase. Ils sont concis par nombre de mots et par longueur: on passe d’un minimum de deux (“En Marche !”, première partie du slogan de Macron et également nom de son mouvement) ou trois mots (“L’avenir en commun”; “Choisir la France”; “Ensemble, la France !”) à neuf mots dans “Le parti qui monte malgré le silence des médias”. En outre, à l’exception de “Nos vies, pas leurs profits !”, qui témoigne d’un “nous” “à récepteur impliqué” (Guespin 1985, p. 48) ou d’un “nous” inclusif large (Labbé 1998, p. 4), aucun slogan n’est directement adressé aux citoyens. On y préfère plutôt le ton de l’exhortation (“En Marche !”; “Ensemble, la France !”) et celui de la dénonciation (“Nos vies, pas leurs profits !”; “Libérez-nous de l’occupation financière !”). La formulation des slogans sous-entend souvent la modalité déontique en tant que nécessité d’agir, de prendre conscience et de s’engager pour viser un objectif commun. Tel est le cas des slogans se commençant par un infinitif ou des slogans averbaux (“Le courage de la vérité”; “Une volonté pour la France”; “L’avenir en commun”; “La force du peuple”; “Au nom du peuple”). Cette tendance apparaît clairement dans les slogans composés de phrases syntaxiquement complètes pourvues d’un verbe modal. Enfin, dans deux cas (si l’on exclut “En Marche !”, hapax des initiales d’Emmanuel Macron), les slogans utilisés sont autoréférentiels et identifient le candidat même qui le porte ou son parti dans le but de marquer une différence par rapport aux autres.

4. Présentation du corpus

Notre corpus de travail est composé des textes des six premières pages de résultats sur Google à partir des seize slogans et du nom du candidat qui le porte: il a été récolté le 26 novembre 2017 et il compte 70575 mots. Les textes recueillis datent de 2016 à 2017 (certains textes vont même au-delà du second tour de l’élection présidentielle): la taille réduite du corpus s’explique par notre choix de ne nous intéresser qu’à une sélection de résultats du web afin de circonscrire l’analyse à ceux qui sont pourvus d’une certaine notoriété, due à leur position au sein du moteur de recherche et à leurs visualisations par les utilisateurs du web. Nous entendons y vérifier les modifications des slogans originels sous la plume des journalistes et des internautes, le poids linguistique et discursif de ces détournements et la possibilité que ceux-ci donnent lieu à des “petites phrases”. La recherche des slogans dans le web est effectuée en tapant le slogan dans sa formulation complète, ou ses mots pleins, y compris éventuellement le nom du candidat concerné si le slogan n’a pas de référence univoque. Tel est le cas d’“Au nom du peuple” et “Le courage de la vérité”, qui indiquent également le début des sentences dans le droit français et le titre des derniers cours de Michel Foucault au Collège de France, respectivement.[7] Pour mieux analyser les slogans, nous avons créé des sous-corpus – un sous-corpus par slogan –, dont la taille et les résultats varient de manière même importante étant donné leur collecte limitée et la recherche, au sein des pages du web, des seuls articles considérés comme pertinents pour notre analyse. Autrement dit, nous avons exclu aussi bien le site du candidat même ou de son parti/mouvement au profit des sites de diffusion d’informations, que l’analyse des réactions du public sous forme de commentaire aux informations publiées. Le Tableau 1 présente les sous-corpus et leur taille:

Titre du sous-corpus	*Taille*
Faire entendre le camp des travailleurs	3998 mots
Le candidat du Frexit	2007 mots
Le parti qui monte malgré le silence des médias	7809 mots
Libérez-nous de l’occupation financière !	12428 mots
Ni système, ni extremes	1046 mots
Le courage de la vérité	5564 mots
Une volonté pour la France	1333 mots
Faire battre le cœur de la France	3671 mots
Je veux retaper la France	13139 mots
Au nom du people	2748 mots
Choisir la France	2178 mots
En Marche !: La France doit être une chance pour tous	5833 mots
Ensemble, la France !	1846 mots
L’Avenir en Commun	1687 mots
La force du people	1386 mots
Nos vies, pas leurs profits !	3902 mots
Nombre total de mots du corpus	70575 mots

Tableau 1
Les sous-corpus.

Nos textes relèvent de la presse, nationale et locale, et des blogs et sites au sein desquels les créateurs et animateurs présentent tant une information officielle qu’une information autre, basée sur l’humour, sur le détournement,

[7] Par conséquent, nous avons exclu de notre analyse les textes du web relevant de ces expressions homonymes des deux slogans présidentiels.

sur les jeux de mots, dont par ailleurs la véracité n’est pas toujours assurée. C’est au sein de ce matériel que nous identifierons les détournements dans la formulation du slogan: les slogans modifiés figurent généralement dans le titre ou comme citations directes ou indirectes rapportées par les journalistes, par les commentateurs, parfois par les politiques. Après leur collecte, les textes seront insérés dans le logiciel *AntConc 3.4.4* (Anthony 2014) dans le but de visualiser les slogans en co(n)texte et d’y extraire, au cas où, de nouveaux slogans potentiels, voire des “petites phrases” par le biais, par exemple, du détournement du slogan originel ou du déplacement de celui-ci du contexte au sein duquel il a été créé à un contexte autre, ainsi engendrant une réception et une interprétation différentes.

Nous sommes convaincue que la notoriété d’un candidat est proportionnelle à la visibilité de son slogan. Or, il en va de même pour la popularité ou l’impopularité du candidat et de son parti/mouvement, mais aussi de son slogan. D’où notre hypothèse, à tester sur le corpus, à partir de laquelle plus un candidat, un slogan, un parti sont impopulaires, plus il est facile que des détournements de son slogan se vérifient, pouvant même donner lieu à des “petites phrases”. Ces détournements sont produits par le web – blogs et réseaux sociaux – et sur le web, qui devient le média et le canal le plus utilisé pour leurs visibilité, diffusion et viralité. Ils sont aptes à permettre aux slogans de poursuivre leur cycle de vie au-delà de la stricte campagne électorale en vertu de la popularité ou impopularité du candidat, du slogan détourné, de sa circulation virale dans le web et de sa médiatisation.

5. Les “petites phrases”

Une “petite phrase” est une expression “extraite des propos d'un personnage politique et abondamment commentée par les médias”[8] dans le but de frapper l'opinion (Brasart 1994). Brasart (1994) fait remonter ce phénomène discursif au discours parlementaire révolutionnaire de la fin du XVIIIe siècle, lorsque, dans leurs discours, les orateurs intègrent la stratégie rhétorique de la réduction à l’égard des dires de leurs adversaires à partir d’une phrase authentiquement proférée ou d’une phrase délibérément déformée. Par une synecdoque, les grands discours sont ainsi réduits à une phrase emblématique qui joue le rôle de porte-parole du tout. En effet, d’un point de vue pragmatique, les “petites phrases” sont des citations pourvues de sens et d’autosuffisance qui se prêtent à une manipulation des utilisateurs et à un détournement par les rapporteurs de ce dire autrui s’il est hors contexte. La diffusion des “petites phrases” créées et rapportées a donc pour but de les faire circuler et de permettre que d’autres politiques se les approprient dans le

[8] *Petit Robert 2017*, version numérique.

temps en créant une sorte de “proverbe post-moderne” (McCallam 2000, p. 56) qui peut être réénoncé. Dans le contexte français, Krieg-Planke et Ollivier-Janiv (2011) remarquent que l’expression “petite phrase” désigne des phénomènes différents: les médias l’emploient comme forme de médiation des discours des responsables politiques. Ceux qui exploitent ces discours confèrent généralement au syntagme une acception négative de déploration, une visibilité (en fait éphémère) et une dégradation du langage politique due à sa diffusion et médiatisation par les responsables politiques et par les médias. Puisqu’elles sont des fragments de discours normalement décontextualisés, les “petites phrases” attirent l’attention et suscitent la polémique (Krieg-Planke, Ollivier-Janiv 2011).

En termes énonciatifs, une “petite phrase” est une phrase détachée, soumise à des contraintes spécifiques (Maingueneau 2011).[9] C’est pourquoi, une fois qu’elle est créée, une “petite phrase” mène une vie propre: détachée de son texte source, elle est découpée et mise en circulation par les médias. Son succès relève des contraintes discursives qui permettent généralement aux énoncés politiquement incorrects de devenir de bons candidats, mais également des intérêts idéologiques qui incitent les acteurs de la communication à disqualifier les propos d’un politique donné. Il en va de même pour notre corpus car les slogans qui se prêtent davantage à être modifiés et à devenir de bons candidats de “petites phrases” sont ceux qui scandalisent et relèvent du politiquement incorrect relativement à la formulation du slogan et à l’idéologie du candidat qui le sous-tend.

D’un point de vue sémantique et formel, les “petites phrases” engendrent un travail sur les formes favorables à la mémorisation et à la reprise, lesquelles peuvent concerner des figures du discours fondées sur des phénomènes sémantiques, tels que la métaphore, et des figures qui relèvent des propriétés formelles des énoncés, telles que la répétition, auxquelles s’ajoutent l’appui sur la doxa et sur les topoï (Krieg-Planke 2011).

Notre définition de “petite phrase”, appliquée à la campagne électorale pour l’élection présidentielle française de 2017, reprend dans les grandes lignes les remarques qui précèdent: il s’agit d’une sorte de slogan détourné par les médias – journalistes et médias comme dispositifs de médiation et de médiatisation (Krieg-Planke 2011) – et par les politiques – les politiques mêmes et leurs communicants. Comme toute “petite phrase”, les slogans qui font l’objet de notre analyse ont une valeur illocutionnaire qui s’explique par l’acte de langage de la promesse et de l’engagement des candidats à l’égard de leurs électeurs potentiels.

[9] Contrairement aux proverbes et aux slogans, les “petites phrases” rentrent dans la logique de la citation car elles portent sur l’extraction d’un fragment de texte (Krieg-Planke 2011).

6. De potentielles "petites phrases" de la présidentielle de 2017?

Par rapport aux slogans politiques analysés, certains candidats et, par conséquent, leurs slogans tendent à capter davantage l'attention des journalistes et des utilisateurs du web et à engendrer des modifications du slogan originel. Tel est le cas des candidats les plus populaires et les plus médiatisés, comme les deux qui arrivent au second tour, Marine Le Pen et Emmanuel Macron, mais également de ceux qui remportent des scores inattendus, dans un sens positif, comme François Asselineau et Jean-Luc Mélenchon, ou négatif, comme Benoît Hamon et François Fillon. C'est pourquoi, suite à la recherche des slogans dans les six premières pages de résultats dans le web, notre analyse a été focalisée sur les candidats les plus aptes à susciter des réactions médiatiques et sur les slogans qui subissent des reprises et/ou des détournements pouvant déboucher sur des "petites phrases" potentielles. Notre examen porte ainsi sur la naissance de la "petite phrase", sur son développement et sur son enracinement au sein d'un contexte donné. En détail, son cycle de vie s'étend sur la campagne pour l'élection présidentielle de 2017 et se conclut avec le scrutin du premier (le 23 avril 2017) et du second tour (le 7 mai 2017).

Pour cette recherche, nous travaillerons sur sept slogans: a) ceux de François Asselineau – candidat-surprise du premier tour et méconnu de la présidentielle de 2017 –, qui définissent tant lui-même que son parti, l'UPR (Union Populaire Républicaine): "Le parti qui monte malgré le silence des médias" et "Le candidat du Frexit"; b) les slogans de François Fillon, vainqueur de la Primaire des Républicains (LR), "Le courage de la vérité" et "Une volonté pour la France"; c) les slogans de Marine Le Pen, présidente du Front National (FN), "Au nom du peuple" et "Choisir la France"; d) le premier des deux slogans du candidat d'En Marche!, Emmanuel Macron, à savoir "En Marche !: la France doit être une chance pour tous".[10] Nous examinerons le traitement de ces slogans par les rapporteurs, les chroniqueurs, les journalistes, les blogueurs, et par les utilisateurs du web qui les commentent jusqu'à les modifier, parfois, pour mettre en relief un trait particulier qui en ressort et qui accroît leur "médiatisation". Nous tenons à préciser que l'appui sur un slogan implique souvent aussi le support sur lequel il apparaît, notamment l'affiche sur laquelle celui-ci est présenté avec, le plus souvent, une photo du candidat. Nos slogans forment ainsi souvent un "iconotexte" (Maingueneau 2015) où les éléments interagissent de manière réciproque.

[10] Ce slogan cumule en fait deux slogans "détachables", dont l'un coïncide avec le nom du mouvement.

6.1. "Le parti qui monte malgré le silence des médias" et "Le candidat du Frexit"

Le slogan de François Asselineau "Le parti qui monte malgré le silence des médias" donne lieu à de nombreux commentaires des journalistes, des commentateurs, des animateurs du web ainsi que du public des internautes du fait des relations souvent difficiles entre ce candidat, qui se considère comme un "bouc émissaire" des médias, et son approche à l'égard des médias. Pourtant, comme également notre corpus en témoigne, c'est cette situation énonciative particulière qui permet à ce politique d'être évoqué, plutôt que par la presse nationale et "officielle", par les titres locaux, par les blogs et par les sites basés sur l'opinion de leurs animateurs, dont le résultat est entre autres une grande visibilité médiatique. C'est ainsi que, à partir de notre recherche des mots pleins du slogan au sein du sous-corpus qui le concerne, composé de 7809 mots, à côté de la formulation exacte du slogan, nous avons relevé des formulations qui peuvent donner lieu à de nouveaux slogans:

(3a) François Asselineau, l'homme qui monte grâce au silence des médias[11]
(3b) Un parti qui croît grâce à Internet[12]
(3c) L'UPR, le parti qui monte malgré le silence des électeurs[13]
(3d) Le parti qui descend malgre [sic] le brouhaha des medias [sic][14]

Ces quatre résultats, qui apparaissent tant entre guillemets, sous forme de citation, que comme titre de texte ou de paragraphe, mettent en évidence une volonté de détourner le slogan originel qui, isolé de son contexte de départ, est modifié pour faire ressortir les spécificités de François Asselineau et de son parti. Les modifications relevées aboutissent à un sens contraire à la direction voulue par le candidat: l'intention est toujours celle de renverser le slogan et de le modifier en termes de montée de consensus, de responsables de ce consensus ou de manque de consensus. C'est par une synecdoque qu'on attribue la caractéristique du parti d'augmenter ses consensus par le silence des médias à son représentant, comme il ressort de (3a). Si (3a) et (3b) n'apparaissent que sous la plume de leurs créateurs, à savoir des journalistes, dans (3c) et (3d) les formulations sont créées par des internautes (le site concerne des jeux vidéo) et leurs occurrences sont uniques. Elles sont

[11] http://www.journal-parabole.com/francois-asselineau-lhomme-qui-monte-malgre-le-silence-des-medias/ (26.11.2017).

[12] http://eleonoredv.over-blog.com/2017/02/traitement-mediatique-de-l-upr-une-anomalie-du-journalisme-politique.html (26.11.2017).

[13] http://www.jeuxvideo.com/forums/42-51-50883151-1-0-1-0-asselineau-l-upr-le-parti-qui-monte-malgre-le-silence-des-medias.htm (26.11.2017).

[14] http://www.jeuxvideo.com/forums/42-51-45718423-1-0-1-0-on-ne-voit-plus-marine-le-pen-depuis-la-branlee.htm (26.11.2017).

caractérisées par la négation du slogan et donnent lieu à des détournements qui représentent pourtant la réalité, puisqu'Asselineau n'obtient que 0,74% de voix au premier tour (le résultat le plus bas après celui de Jacques Cheminade, avec 0,18% de voix).

Les détournements de ce slogan diffèrent de ceux qui émergent de l'analyse du second slogan d'Asselineau, à savoir "Le candidat du Frexit". Notre sous-corpus, composé de 2007 mots, donne les résultats suivants:

(2a) Je suis le candidat du Frexit[15]
(2b) Je suis le seul candidat du Frexit[16]
(2c) François Asselineau candidat du Frexit progressiste ? pour le moins on peut en douter ![17]
(2d) Asselineau candidat du Frexit ? il faut en douter. Candidat pro système ? c'est certain[18]
(2e) [L]e voyage au bout du Frexit du candidat Asselineau[19]
(2f) François Asselineau, le candidat anti-européen et anti-américain[20]

Ces formulations sont tirées de titres d'analyses et d'articles de presse, ainsi que de citations de François Asselineau ((2a) et (2b)): elles visent à sceller le lien entre le candidat et sa conception du "Frexit". En témoignent les appositions et les épithètes qui accompagnent le slogan originel et qui insistent sur le crédo de ce candidat opposé au système politique et économique français de son époque (2f). Cela est à notre avis également perceptible depuis le titre de (2e), là où la reformulation du slogan originel par "Le voyage au bout du Frexit du candidat Asselineau" rappelle le titre du roman de Céline "Le voyage au bout de la nuit", qui se caractérise par le fait de combattre le système existant à l'époque de sa rédaction, la guerre, le capitalisme, le colonialisme. Ce qui en ressort, c'est également un candidat qui se distingue des autres par l'exclusivité des propos énoncés: tel est le cas de son épithète "seul", dans ses propres mots ((2g) et (2h)), synonyme de "vrai" et opposé "aux autres":

[15] https://www.francetvinfo.fr/politique/francois-asselineau/presidentielle-je-suis-le-candidat-du-frexit-soutient-francois-asselineau_2118837.html (26.11.2017).
[16] http://www.ledauphine.com/politique/2017/04/11/je-suis-le-seul-candidat-du-frexit (26.11.2017).
[17] https://www.initiative-communiste.fr/articles/europe-capital/asselineau-candidat-du-frexit-progressiste-pour-le-moins-on-peut-en-douter/ (26.11.2017).
[18] https://www.initiative-communiste.fr/articles/europe-capital/asselineau-candidat-du-frexit-progressiste-pour-le-moins-on-peut-en-douter/ (26.11.2017).
[19] http://www.lesinrocks.com/2017/04/20/actualite/politique/nantes-le-voyage-au-bout-du-frexit-du-candidat-asselineau-11935758/ (26.11.2017).
[20] https://start.lesechos.fr/actu-entreprises/societe/petitscandidats/francois-asselineau-le-candidat-anti-europeen-et-anti-americain-7934.php (26.11.2017).

> (2g) "Et je suis le seul à proposer de sortir de l'UE, le seul candidat du Frexit, pas comme Mme Le Pen qui veut seulement renégocier les traités !"[21]
>
> (2h) "Il y a un seul candidat pour le Frexit, et les dix autres pour dire de belles paroles, ce sont des menteurs !"[22]

Les intentions du candidat sont pourtant parfois détournées par la médiatisation de cette citation, puisque ceux qui la rapportent le font toujours en la référant aux mots prononcés par Asselineau, entre guillemets, plutôt que comme résultat "objectif" de l'analyse de son programme.

6.2. "Le courage de la vérité" et "Une volonté pour la France"

Les slogans de François Fillon, candidat au premier tour et représentant des Républicains, sont ceux qui ont fait couler le plus d'encre pendant la campagne pour l'élection présidentielle de 2017. "Le courage de la vérité" apparaît comme un plagiat politique du slogan d'Arnaud Montebourg pour les élections législatives de 1997, 2002 et 2007, mais il coïncide également et surtout avec le titre des derniers Cours donnés au Collège de France par Michel Foucault en 1983-1984. Cette citation d'envergure acquiert d'autant plus d'importance si on la rapporte à la campagne électorale de François Fillon, considéré, même avant son début de campagne, comme l'un des favoris. Son slogan est suivi de "c'est la condition pour redresser notre pays. Seule la vérité peut nous permettre de regarder avec lucidité les difficultés du pays et d'y apporter des réponses efficaces. La vérité est la clé du redressement national". Le sens du slogan et la citation de Fillon mettent en évidence que la vérité est la seule possibilité de redresser la France.

Une grande campagne médiatique, initiée par des déclarations des journalistes du *Canard Enchaîné*, aboutit, dès janvier 2017, à la mise en examen de ce candidat pour cause de présumés emplois fictifs qu'il aurait accordés à sa femme et à ses enfants. Depuis ces déclarations, le devant de la scène politico-médiatique est occupé par le lien entre le slogan utilisé par Fillon, invoquant "le courage de la vérité", et les événements judiciaires qui mettent en cause sa personne et son personnage politique. La recherche des mots pleins de ce slogan dans notre corpus confirme ce rapprochement et donne lieu à de nombreux détournements, alors que le sens originel du slogan et les déclarations du candidat ne figurent désormais que dans des vidéos de YouTube. Face à cette affaire, pour laquelle le candidat se déclare innocent, il décide de ne pas renoncer à sa course pour la Primaire de la droite et pour l'élection présidentielle, et de modifier son slogan. Dans le nouveau slogan,

[21] http://www.lalsace.fr/actualite/2017/04/12/je-suis-le-seul-candidat-du-frexit (26.11.2017).
[22] https://presidentiellesenportraits.com/2017/04/11/francois-asselineau-le-candidat-du-frexit-attire-1500-personnes-a-lille/ (26.11.2017).

paru le 19 mars 2017, "Une volonté pour la France", les mots "courage" et "vérité" disparaissent. Le web, dans la presse mais aussi et surtout *via* les blogs et les réseaux sociaux, rapportés par la presse en ligne, fourmille en commentaires, pour la plupart négatifs, à propos du slogan originel de François Fillon et de son second slogan créé à la hâte. Par rapport au premier slogan, dans notre sous-corpus, de 5564 mots, les formulations isolées jouent surtout sur la forme et sur le sens de "courage" et "vérité" à l'issue de l'affaire "Penelopegate" à partir des déclarations du candidat, avant que les accusations à son égard ne soient manifestées:

> (6a) François Fillon abandonne "le courage de la vérité"[23]
> (6b) Il abandonne "le courage de la vérité". Exit le courage. Exit la vérité ![24]
> (6c) François Fillon face à Michel Foucault : "le courage de la vérité"[25]
> (6d) Au revoir "courage" et "vérité"[26]
> (6e) François Fillon, le courage de la post-vérité[27]
> (6f) Sur l'état de la France, François Fillon avait le courage de la vérité[28]
> (6g) "Courage Fillon !"[29]

En particulier, (6e) exprime l'incongruité des acceptions de "courage" et "vérité" auprès de Michel Foucault, lesquelles, selon les commentateurs, ne se retrouvent pas dans l'attitude de François Fillon: "vérité" y est remplacé par "post-vérité", néologisme indiquant que "les discours démagogiques et les fausses informations influencent davantage l'opinion que les faits objectifs".[30] Au contraire, le doute et la présumée innocence de Fillon résultent de (6f) et (6g), dont ce dernier – il s'agit par ailleurs d'une paronomase du titre du film français de 1979 "Courage fuyons"[31] à partir des deux paronymes "Fillon" et "fuyons" –, sonne comme une exhortation, sous-tendue au mot "courage", à poursuivre la recherche de la vérité. Des formulations de ce type sont pourtant uniques dans notre sous-corpus.

[23] https://blogs.mediapart.fr/jchouelfreefr/blog/200317/francois-fillon-abandonne-le-courage-de-la-verite (26.11.2017).
[24] https://blogs.mediapart.fr/jchouelfreefr/blog/200317/francois-fillon-abandonne-le-courage-de-la-verite (26.11.2017).
[25] https://www.franceculture.fr/emissions/petit-precis-de-philosophie-lusage-des-candidats/francois-fillon-face-michel-foucault-le (26.11.2017).
[26] http://www.20minutes.fr/high-tech/2033655-20170319-francois-fillon-supprime-mot-verite-slogan-campagne-suscite-moqueries-twitter (26.11.2017).
[27] http://jam.unine.ch/index.php/2017/05/23/post-verite-a-modifier/ (26.11.2017).
[28] http://www.lepolitique.fr/2017/06/30/sur-letat-de-la-france-francois-fillon-avait-le-courage-de-la-verite/ (26.11.2017).
[29] http://www.francisrichard.net/2017/02/courage-fillon.html (26.11.2017).
[30] Article "post-vérité", *Petit Robert 2017*, version numérique.
[31] Nous tenons à remercier l'un de nos relecteurs anonymes pour cette intéressante suggestion.

En revanche, les articles de presse se sont engagés à donner une grande visibilité à plusieurs slogans détournés, créés par les internautes et diffusés à travers les réseaux sociaux, à partir du second slogan de François Fillon, “Une volonté pour la France”. Si les mots “volonté” et “France” sont susceptibles de subir de légères modifications graphiques et phonétiques, leur sens est complètement détourné et reflète le climat qui fait suite aux accusations contre François Fillon:

(7a) “Un vol pour la France”[32]
(7b) “Un voleur pour la France”[33]
(7c) “Un vol éhonté pour la France”[34]
(7d) “Une volonté pour la fraude”[35]

6.3. “Au nom du peuple” et “Choisir la France”

Les autres slogans analysés concernent les deux candidats qui remportent le premier tour, à savoir Emmanuel Macron et Marine Le Pen. Dans les deux cas, nous avons affaire à deux slogans, dont le dernier, créé pour le second tour, est chronologiquement second par rapport aux slogans de campagne premiers, qui accompagnent les candidats pendant la plupart de leur campagne, ont un cycle de vie plus long et sont ainsi sujets à plus de modifications et de détournements.

Le slogan mariniste “Au nom du peuple” a des origines anciennes: il apparaît dans l’en-tête des jugements, notamment sous la formule (Krieg-Planke 2009) “Au nom du peuple français”, raccourci par Marine Le Pen en éliminant l’épithète.[36] Pourtant, avant de devenir un slogan de campagne, “Au nom du peuple” est le nom d’une association lyonnaise apolitique protectrice des droits de l’homme, qui ne veut pas être assimilée à Marine Le Pen. C’est cette dernière qui obtient néanmoins gain de cause et peut donc utiliser “Au nom du peuple” comme slogan pour sa campagne électorale face à l’association, qui avait mis le parti en demeure. Notre sous-corpus, composé de 2748 mots, montre cette querelle sur la paternité du slogan et questionne ainsi le choix du slogan, auquel on lie également celui de Jean-Marie Le Pen “Le Pen, le peuple”, que Marine Le Pen modifie en supprimant toute

[32] https://francais.rt.com/france/35968-nouvelle-affiche-francois-fillon-deja-objet-detournements (26.11.2017).
[33] https://francais.rt.com/france/35968-nouvelle-affiche-francois-fillon-deja-objet-detournements (26.11.2017).
[34] https://francais.rt.com/france/35968-nouvelle-affiche-francois-fillon-deja-objet-detournements (26.11.2017).
[35] https://francais.rt.com/france/35968-nouvelle-affiche-francois-fillon-deja-objet-detournements (26.11.2017).
[36] https://www.lci.fr/elections/d-ou-vient-au-nom-du-peuple-le-slogan-de-campagne-de-marine-le-pen-2003636.html (26.11.2017).

référence à son patronyme et à son père. Lorsque les journalistes et les internautes modifient le slogan, ils le font pour démentir le renvoi au peuple, présumé faux, en enrichissant le slogan d'éléments qui l'éloignent du peuple. C'est le cas de:

(10a) La "candidate de la France du peuple"[37]
(10b) "Candidate du peuple"[38]
(10c) "C'est en votre nom que je prends la parole"[39]
(10d) "Tout au nom du peuple français. Rien sans lui, rien contre lui"[40]
(10e) "C'est en candidate du peuple que je me présente devant vous"[41]
(10f) Le "Système", c'est le "déni du peuple"[42]
(10g) Marine Le Pen, une candidate du "peuple" face aux partis du "fric"[43]
(10h) Marine Le Pen, une "patriote" contre les "mondialistes"[44]
(10i) "Au nom du peuple"… et du capital[45]

(10a), (10b), (10c), (10d), (10e), (10f) se réfèrent à des déclarations prononcées par Marine Le Pen lors de ses discours de campagne, que les journalistes isolent comme titre de texte ou de paragraphe et emploient de manière "neutre". Dans (10g), en revanche, les journalistes reformulent sa déclaration "Contre la droite du fric, la gauche du fric, je suis la candidate de la France du peuple" en mettant en relief les deux mots "peuple" et "fric", opposés du point de vue aussi bien de registre de langue ("fric" relève de la langue familière) que surtout sémantique pour sceller la distance et la différence entre la droite et la gauche ("les partis du fric") et elle-même, "candidate de la France du peuple". Une reformulation touche également à (10h), où Marine Le Pen est qualifiée de "patriote" – par rapport également à son tract de campagne "Les patriotes au pouvoir" – en opposition avec les "mondialistes", à savoir ses adversaires politiques. Dans (10i), enfin, c'est le Nouveau parti anticapitaliste qui contredit le slogan originel de Marine Le

[37] https://selp.eu/figures/je-suis-la-candidate-de-la-france-du-peuple/ (26.11.2017).
[38] https://www.lci.fr/elections/video-resultats-presidentielle-2017-je-suis-la-candidate-du-peuple-le-discours-de-marine-le-pen-apres-l-annonce-des-resultats-provisoires-2045979.html (26.11.2017).
[39] https://www.nouvelobs.com/politique/election-presidentielle-2017/20160919.OBS8333/marine-le-pen-change-de-slogan-une-ligne-populiste-assumee.html (26.11.2017).
[40] http://www.bfmtv.com/politique/au-nom-du-peuple-une-association-conteste-le-slogan-de-campagne-du-front-national-1039281.html (26.11.2017).
[41] http://www.lepoint.fr/politique/2017-marine-le-pen-veut-se-presenter-en-candidate-du-peuple-18-09-2016-2069382_20.php (26.11.2017).
[42] https://francais.rt.com/france/31744-systeme-deni-peuple-marine-le-pen-campagne (26.11.2017).
[43] http://www.lepoint.fr/presidentielle/marine-le-pen-une-candidate-du-peuple-face-aux-partis-du-fric-05-02-2017-2102580_3121.php (26.11.2017).
[44] https://www.ouest-france.fr/politique/marine-le-pen/marine-le-pen-une-patriote-contre-les-mondialistes-4782660 (26.11.2017).
[45] https://npa2009.org/actualite/politique/le-pen-au-nom-du-peuple-et-du-capital (26.11.2017).

Pen par le biais d'une expansion, en soulignant le lien entre le populisme de Marine Le Pen et son vrai attachement au capitalisme – là où le FN se réclame anticapitaliste.

Le second slogan de Marine le Pen, "Choisir la France", est accompagné d'une photo apparemment retouchée de Marine Le Pen qui figure sur les affiches de campagne.[46] Dans notre sous-corpus, composé de 2178 mots, de nombreux détournements sont rapportés par la presse en ligne et par les bloggeurs afin de démystifier l'image irréelle de Marine Le Pen et le sens du slogan originel. Celui-ci relève de l'exclusivité par rapport au reste du monde (ainsi contre l'immigration, contre la mondialisation, contre l'Europe et le monde entier, comme le prônent les discours et le programme de Marine Le Pen et du FN). C'est le cas de:

(9a) "Choisir la France ? Choisir Photoshop oui !"[47]
(9b) Choisir Photoshop[48]
(9c) Choisir la frange[49]
(9d) Choisir Air France Retour gratuit au pays[50]
(9e) Choisir la race[51]
(9f) Moisir la France[52]
(9g) Moisir la rance[53]
(9h) Choisir la francisque[54]
(9i) Choisir la France, c'est renoncer à Marine Le Pen[55]
(9j) "Entre la finance et la France, il faut choisir"[56]

Les détournements comportent souvent le changement d'un ou de deux mots du slogan de départ ou plutôt la modification et/ou l'élimination de quelques

[46] Le web fourmille dans des commentaires qui attribuent le corps de l'affiche à Marion Maréchal-Le Pen, nièce de Marine Le Pen.
[47] http://www.20min.ch/ro/news/insolite/story/-Choisir-la-France--Choisir-Photoshop--oui---13995729 (26.11.2017).
[48] https://www.lexpress.fr/actualite/politique/elections/moisir-la-france-l-affiche-de-campagne-de-marine-le-pen-detournee_1902645.html (26.11.2017).
[49] https://www.lexpress.fr/actualite/politique/elections/moisir-la-france-l-affiche-de-campagne-de-marine-le-pen-detournee_1902645.html (26.11.2017).
[50] https://www.lexpress.fr/actualite/politique/elections/moisir-la-france-l-affiche-de-campagne-de-marine-le-pen-detournee_1902645.html (26.11.2017).
[51] https://www.lexpress.fr/actualite/politique/elections/moisir-la-france-l-affiche-de-campagne-de-marine-le-pen-detournee_1902645.html (26.11.2017).
[52] https://www.lexpress.fr/actualite/politique/elections/moisir-la-france-l-affiche-de-campagne-de-marine-le-pen-detournee_1902645.html (26.11.2017).
[53] https://www.lexpress.fr/actualite/politique/elections/moisir-la-france-l-affiche-de-campagne-de-marine-le-pen-detournee_1902645.html (26.11.2017).
[54] https://www.tuxboard.com/affiche-campagne-marine-le-pen/ (26.11.2017).
[55] http://www.atlantico.fr/decryptage/choisir-france-c-est-renoncer-marine-pen-3039423.html (26.11.2017).
[56] https://www.publicsenat.fr/article/politique/nicolas-dupont-aignan-au-meeting-de-marine-le-pen-entre-la-finance-et-la-france-il (26.11.2017).

lettres pour lui conférer des interprétations très différentes. Si, dans (9a), (9b) et (9c), le changement de COD touche à un aspect purement visuel et aux retouches de la photo et, dans le cas de "France"/ "frange", il peut susciter le sourire, les autres formulations sous-tendent, au-delà du jeu de mots, un travail plus subtil sur le crédo de Marine Le Pen. Ainsi (9d) fait-il référence à la fermeture des frontières et à l'immigration zéro, mesure-phare du programme électoral de la candidate, tout comme (9e), qui voit le remplacement de "France" par "race", ce qui pousse un peu plus loin le discours mariniste. Au contraire, les détournements de (9f) et (9g) portent sur "moisir" et "rance", lesquels soulignent une détérioration de la France, notamment celle que, selon leurs créateurs, la France traverserait si Marine Le Pen devenait Présidente de la République. Enfin, le détournement de (9h) renvoie, par le remplacement de "France" par "francisque", à la hache de guerre des Francs devenue symbole du régime de Vichy, parfois même évoquée par des élus du FN sous sa devise "Travail, Famille, Patrie". Il s'agit de détournements pouvant susciter le sourire, mais ces formulations sont potentiellement possibles et aisées à mémoriser et à retenir en raison de leur médiatisation par la presse et par les blogs en ligne et du poids de l'image accompagnant le slogan dans les affiches. Enfin, seuls les deux derniers slogans portent sur le contenu du slogan originel: (9i) comporte la négation du slogan, alors que (9j) oppose les deux alternatives entre lesquelles Marine Le Pen devrait choisir: la France ou la finance.

6.4. "En Marche ! : La France doit être une chance pour tous"

Enfin, nous n'avons retenu que la première partie du premier slogan d'Emmanuel Macron, "En Marche ! : la France doit être une chance pour tous", alors que l'attention médiatique envers son slogan de l'entre-deux-tours, "Ensemble, la France", est beaucoup moins élevée au profit du slogan de Marine Le Pen ci-dessus.

Notre sous-corpus se compose de 5833 mots: le slogan y figure sous forme de titre de presse, avec des reformulations, des précisions ou des détournements. Ces derniers concernent, encore une fois, les mots et les lettres qui les composent au sein du slogan.

(12a) Emmanuel Macron, une chance pour les mantais ?[57]
(12b) Macron pourrait être une chance pour la Nouvelle-Calédonie[58]

[57] http://plumesdemantes.com/index.php/2017/05/07/emmanuel-macron-une-chance-pour-les-mantais/ (26.11.2017).
[58] http://www.dnc.nc/patrick-louis-macron-pourrait-etre-une-chance-pour-la-nouvelle-caledonie/ (26.11.2017).

(12c) La France numérique doit être une chance pour tous[59]
(12d) Pour Macron, l'immigration "se révèle une chance" pour la France[60]
(12e) L'Allemagne voit en Emmanuel Macron une chance pour l'Europe[61]
(12f) Tout le monde doit être une chance pour la France[62]
(12g) Une chance pour tous ? Et que fait-on de nos 1500 ans d'histoire purement blanche ?[63]
(12h) Avec Macron, ce sera l'islamisme en marche ! Le communautarisme en marche ![64]
(12i) Macron, "Une chance pour tous" ? Mais sûrement pas pour la France[65]
(12j) Macron, c'est l'anti-France "En Marche" ![66]
(12k) Programme Macron: Une menace pour la France ![67]
(12l) La "chance pour tous"...un slogan piqué au Loto ?[68]

Les premières formulations touchent au mot "chance" et à ses expansions. Nous avons affaire à des titres d'articles qui présentent Emmanuel Macron comme une "chance" pour ceux qui viennent de certaines parties de France (tel est le cas de Mantes-la-Joie (12a) et de l'Outre-Mer (12b)) ou bien à un aspect de France potentiellement riche en opportunités, comme le numérique (12c). D'autres interprétations supposant un partage du programme de Macron concernent son attitude favorable à l'égard de l'immigration (12d), sur laquelle mise entre autres l'Allemagne (12e). Au contraire, au-delà de (12f), où l'inversion des éléments du slogan donne lieu à un non-sens, les autres détournements sous-tendent un désaccord sur le programme du candidat d'En Marche! en matière d'immigration (12g), avec également le risque d'une montée des communautarismes, voire de l'islamisme (12h), ou bien une opposition totale à ce candidat. Tel est le cas de (12i), (12j), (12k), pour la plupart tirés des déclarations de Marine Le Pen, là où le programme

[59] https://medium.com/je-vote-macron/la-france-num%C3%A9rique-doit-%C3%AAtre-une-chance-pour-tous-1f18e6aef6f1 (26.11.2017).
[60] http://www.valeursactuelles.com/politique/pour-macron-limmigration-se-revele-une-chance-pour-la-france-70488 (26.11.2017).
[61] https://www.la-croix.com/Monde/Europe/LAllemagne-voit-Emmanuel-Macron-chance-pour-lEurope-2017-04-25-1200842386 (26.11.2017).
[62] http://www.jeuxvideo.com/forums/42-51-50547384-1-0-1-0-macron-une-chance-pour-la-france.htm (26.11.2017).
[63] http://www.jeuxvideo.com/forums/42-51-50547384-1-0-1-0-macron-une-chance-pour-la-france.htm (26.11.2017).
[64] http://www.europe1.fr/politique/marine-le-pen-avec-macron-ca-sera-lislamisme-en-marche-le-communautarisme-en-marche-3295730 (26.11.2017).
[65] http://lagauchematuer.fr/2017/04/28/emmanuel-macron-ce-sera-leonarda-en-pire/ (26.11.2017).
[66] http://www.valeursactuelles.com/politique/pour-macron-limmigration-se-revele-une-chance-pour-la-france-70488 (26.11.2017).
[67] http://www.lemonde.fr/election-presidentielle-2017/article/2017/03/02/programme-macron-une-menace-pour-la-france-selon-marine-le-pen_5088429_4854003.html (26.11.2017).
[68] http://speech.konbini.com/news/le-grand-nimporte-quoi-des-reseaux-sociaux-special-affiche-et-slogan-de-macron/ (26.11.2017).

de Macron est considéré comme une menace pour la France. Enfin, les internautes s’attachent à la formulation même du slogan originel et à l’acception du mot “chance” comme “fortune/ sort” (12l) plutôt que comme “possibilité” et/ou “opportunité”.

7. Remarques

Une “petite phrase” est souvent le résultat d’un détournement, d’une déploration du discours politique par les médias, qui modifient délibérément une déclaration d’un politique ou de son cabinet pour diffuser et médiatiser ces nouveaux contenus. Dans les slogans examinés, la forme prime souvent sur le fond et les slogans originels sont isolés de leur contexte pour être réutilisés différemment tant dans leur aspect lexical que discursif. Un trait essentiel qui émerge des détournements de nos slogans et des “petites phrases” est leur événementialité et adaptabilité difficile en dehors du contexte de leur création. Si certains détournements relèvent de l’humour, d’autres portent sur des accusations, voire sur des tons et sur des langages qui offensent le candidat concerné. En particulier, il est possible d’identifier deux détournements principaux. Le premier n’est pas un détournement à proprement parler, mais une reformulation ou une expansion du slogan originel par son “détenteur” ou par les commentateurs et les internautes qui le proposent: tel est le cas des déclarations de François Asselineau et de celles de Marine Le Pen à propos de leurs propres slogans, et de celles des journalistes qui commentent de manière “neutre” les propos présentés dans les slogans des quatre candidats examinés. Ces modifications du slogan originel sont à notre avis les moins susceptibles à devenir une “petite phrase”. Pour le second détournement, on peut distinguer deux cas: d’une part, les slogans originels sont modifiés dans leur forme par le biais surtout de paronymes en vue de créer des effets sonores et graphiques particuliers suscitant le sourire ou la blague, là où le sens du slogan originel est modifié ou annulé. A ce propos, nous avons relevé que même les éléments non langagiers, dont les images sous forme de photo – si l’on pense au slogan de l’entre-deux-tours de Marine Le Pen –, ont leur rôle à jouer dans la déformation du slogan. D’autre part, le détournement du slogan relève de sa négation à partir d’un crédo politique opposé à celui du candidat du slogan. Cela se produit lorsque des adversaires politiques, des commentateurs et des internautes engagés jouent sur les mots du slogan de manière à créer la polémique. Nous estimons que cette dernière catégorie est la plus apte à l’émergence de “petites phrases” puisque les modifications des slogans sont réalisées en vue de frapper. En revanche, un détournement “faible” qui ne suscite pas la polémique n’engendrerait pas de “petites phrases”. Les détournements des deux slogans de François Fillon, y compris le slogan même “Le courage de la vérité” suite à l’affaire judiciaire du candidat, ainsi que les slogans de Marine Le Pen, notamment les détournements à des fins politiques de “Choisir la France”, peuvent donc devenir des “petites phrases”. Quant aux slogans de François Asselineau, c’est d’abord la longueur de son premier slogan originel (de neuf mots) qui se prête moins à l’émergence de “petites phrases”, au-delà des détournements produits. Enfin, le slogan macronien nous semble être, à l’exception de ses détournements par Marine Le Pen, l’expression de reformulations et d’expansions plutôt que de potentielles “petites phrases”, d’autant plus que le ton utilisé est souvent celui de l’espoir – inscrit dans les acceptions mêmes de “chance” – plutôt que de la polémique et de l’accusation.

8. Conclusions

Dans les slogans politiques de la campagne présidentielle de 2017 en France, nous avons identifié les phases de leur cycle de vie: leur création, leur diffusion et leurs éventuels détournements. C’est notamment cette dernière phase que nous avons essayé d’analyser car elle peut coïncider avec la construction d’une “petite phrase” en politique, c’est-à-dire avant que “[c]ette fraction d’un tout, parce qu’elle s’avère particulièrement frappante, étonnante ou polémique parvien[ne] à se décontextualiser, s’autotomiser, en quelque sorte, et se met[te] à circuler dans l’interdiscours médiatique” (Seoane 2017, p. 3).

Les phénomènes observés au cours de cette recherche sur les détournements des slogans politiques sont ainsi susceptibles à déboucher sur des “petites phrases”. Nous avons exploité un petit corpus tiré de résultats ciblés issus du web à partir des mots pleins de chaque slogan afin de déceler les modifications éventuelles qu’il peut subir sous la plume de ceux qui le rapportent sur la Toile à des fins de visibilité mais surtout d’opposition et de dénonciation. Or, bien que la campagne présidentielle de 2017 soit arrivée à son terme avec le second tour de l’élection présidentielle, ses retombées et les analyses politiques et scientifiques auxquelles elle donne lieu sont allées au-delà de la date des élections. Si la plupart des détournements présentés sont à notre avis potentiellement aptes à devenir des “petites phrases” en termes formels (ils sont brefs, ils captent l’attention, ils sont faciles à mémoriser), d’un point de vue sémantique ce sont ceux qui frappent de manière négative qui peuvent acquérir le statut de “petite phrase”. Nous excluons ainsi les détournements des éléments graphiques des slogans aux seules fins de mémorisation ou d’amusement, alors que nous retenons ceux qui, provenant surtout des journalistes et des opposants politiques, attirent l’attention des lecteurs en termes souvent polémiques.

L'analyse des slogans détournés de la présidentielle de 2017 nous permet donc de confirmer qu'une "petite phrase" naît et se développe dans le champ politique et ensuite médiatique; elle résulte d'un discours politique volontairement fragmenté en vue de sa médiatisation, qui se charge ainsi d'une forte valeur illocutionnaire (Krieg-Planke 2011). Elle est connotée négativement et fait référence à un énoncé qui est dénié, critiqué, dénoncé, déprécié, condamné; elle est associée à la surface et à la forme par rapport au fond, mais aussi à l'anecdote par rapport au contenu, à la polémique, à l'injure, à l'irresponsabilité et à l'annonce en opposition avec l'action (Krieg-Planke 2011). Dans les cas présentés, nous avons souvent affaire à une extension (épithètes, appositions) du slogan originel qui, isolé de son contexte, est réutilisé différemment dans son aspect lexical et discursif. De plus, nos exemples de détournements montrent que les formulations créées à partir des slogans originels relèvent d'un langage non-savant et d'acteurs qui tendent à contredire et à détourner le dire d'autrui plutôt qu'à l'appuyer et à s'y reconnaître.

Pour conclure, ce travail n'a porté ni sur la distinction scientifique rigoureuse entre "slogan" et "petite phrase" ni sur leur plus ou moins stricte filiation – ce sont autant de thèmes qui demanderaient des analyses *ad hoc* –, mais sur les trajectoires discursives qu'un slogan politique tiré d'un événement politique majeur, notamment l'élection présidentielle en France, peut prendre sous la plume de ceux qui en détournent la forme et le contenu. Parmi ces trajectoires, nous nous sommes intéressée à celle qui aboutit à une "petite phrase", notamment en présentant, en amont, les étapes lexicales, pragmatiques et discursives par lesquelles les modifications peuvent transformer le slogan en "petite phrase" de la politique française. C'est pourquoi, au bout de cette recherche, il nous a été possible d'identifier, parmi les détournements de slogan, des "petites phrases potentielles", pour reprendre notre titre. Néanmoins, afin que ces dernières formulations deviennent des "petites phrases", elles ont besoin de mener leur cycle de vie dans le temps, notamment au-delà de l'espace de la stricte campagne présidentielle ainsi que du temps et de l'espace de la collecte de notre corpus.

Note biographique: Alida M. Silletti est enseignant-chercheur en Langue et traduction françaises auprès de l'Université de Bari Aldo Moro (Italie) depuis 2012. Elle a obtenu un doctorat en Langues et cultures comparées auprès de l'Université de Modène et Reggio d'Emilie (Italie) en 2009. Ses travaux de recherche portent sur la linguistique du verbe comparée français-italien et sur la morphosyntaxe de la phrase complexe française et italienne. Elle effectue également des recherches sur le discours de vulgarisation scientifique et juridique français et italien, sur le discours de représentation sur l'Europe, sur le discours hypertextuel des médias et sur le discours politique et identitaire français.

Adresse électronique: alida.silletti@uniba.it

Références

Adam J.-M. 2008, *La linguistique textuelle. Introduction à l'analyse textuelle des discours*, A. Colin, Paris.

Amossy R. 2002, *Ethos*, in Charaudeau P., Maingueneau D. (ed.), *Dictionnaire d'analyse du discours*, Seuil, Paris, pp. 238-240.

Anthony L. 2014, *AntConc 3.4.4*. http://www.laurenceanthony.net/software/antconc/ (31.08.2018).

Authier-Revuz J. 1996, *Remarques sur la catégorie de l'"îlot textuel"*, in "Cahiers du français contemporain" 3, numéro thématique *Hétérogénéités en discours*, pp. 91-116.

Barbet D. 2012, *Les emprunts discursifs entre politique et publicité. Des échanges inégaux*, in "Mots. Les langages du politique" 98. http://mots.revues.org/20590 (24.02.2018).

Brasart P. 1994, *Petites phrases et grands discours (Sur quelques problèmes de l'écoute du genre délibératif sous la Révolution française)*, in "Mots. Les langages du politique" 40, numéro spécial *Ecoutes, échos du politique*, pp. 106-112.

Charaudeau P. 2005, *Le discours politique. Les masques du pouvoir*, Vuibert, Paris.

Charaudeau P. 2009, *Il n'y a pas de société sans discours propagandiste*, in Ollivier-Yaniv C., Rinn M. (ed.), *Communication de l'État et gouvernement social,* PUG, Grenoble, pp. 19-38.

Charaudeau P. 2015, *Le charisme comme condition du leadership politique*, in "Revue française des sciences de l'information et de la communication" 7. http://journals.openedition.org/rfsic/1597 (31.08.2018).

Guespin L. 1985, Nous*, la langue et l'interaction*, in "Mots. Les langages du politique" 10, numéro spécial *Le nous politique*, pp. 45-62.

Jaubert M.-J. 1985, *Slogan mon amour*, Barrault Bernard, Paris.

Krieg-Planke A. 2009, *La notion de "formule" en analyse du discours. Cadre théorique et méthodologique*, Presses universitaires de Franche-Comté, Besançon.

Krieg-Planke A. 2011, *Les "petites phrases": un objet pour l'analyse des discours politiques et médiatiques*, in "Communication & langages" 168, pp. 23-41.

Krieg-Planke A., Ollivier-Yaniv C. 2011, *Poser les "petites phrases" comme objet d'étude*, in "Communication & langages" 168, pp. 17-22.

Labbé D. 1998, *Le "nous" du général de Gaulle*, in "Quaderni di studi linguistici" 4/5, pp. 331-354.

Maingueneau D. 2002, *Problèmes d'ethos*, in "Pratiques" 113-114, pp. 55-68.

Maingueneau D. 2011, *Sur une petite phrase "de" Nicolas Sarkozy. Aphorisation et auctorialité*, in "Communication et langages" 188, pp. 43-56.

Maingueneau D. 2015, *Argumentation et scénographie*, in Pineira-Tresmontant C. (ed.), *Discours et effets de sens. Argumenter, manipuler, traduire*, Artois Presses Université, Arras, pp. 71-86.

Mayaffre D. 2003, *Dire son identité politique: Etude du discours politique français français au XXème siècle*, in "Cahiers de la Méditerranée", Centre de la Méditerranée Moderne et Contemporaine (CMMC) - Université de Nice-Sophia Antipolis, Nice, pp. 247-264.

McCallam D. 2000, *Les "petites phrases" dans la politique anglo-saxonne*, in "Communication & langages" 126, pp. 52-59.

Nianxi X. 2008, *Logique et slogans en politique*, in "Diogène" 221 [1], pp. 146-155.

Reboul O. 1975, *Le slogan*, Éditions Complexe, Bruxelles.

Rey A. 2017, *Le Petit Robert de la langue française 2017* (version numérique).

Seoane A. 2017, *Petite phrase*, in *Publictionnaire. Dictionnaire encyclopédique et critique des publics*. http://publictionnaire.huma-num.fr/notice/petite-phrase/ (31.08.2018).

PART II Popularisation media

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 339-358
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p339
http://siba-ese.unisalento.it,

DISSEMINATING AND ADAPTING SPECIALIZED KNOWLEDGE
American think tanks' blogs

MATHILDE GAILLARD
UNIVERSITÉ PARIS SORBONNE

Abstract – The present study aims to characterize the ways in which specialized knowledge is disseminated and adapted in the discourse of American think tanks whose specificity lies in their objective to influence public policy. It offers a comparative rhetorical analysis of an exploratory corpus composed of blog posts and their corresponding expert reports published by six think tanks between 2014 and 2017. The following hypothesis is explored: blog posts are characterized by distinctive rhetorical strategies in the discursive space of think tanks and may thus be seen as a means for these organizations to carry out their programmatic aim. Results show that the format of the genre, as evidenced by the use of journalistic techniques such as titles and hooks, is particularly tailored to meet the needs and draw the attention of a wide audience on experts' work. An analysis of argumentative choices and hedging in the corpus further suggests that blogs may also represent a way for think tank experts to position themselves in the political arena. These overlapping communicative purposes more generally testify to the specialized nature of a new genre in think tanks' outreach strategy.

Keywords: blogs; think tanks; United States; programmatic aim; rhetorical strategies; genre analysis.

1. Introduction

Think tanks are one of the many actors involved in the decision-making process in modern-day US society, as their specificity lies in their desire to influence public policy by disseminating concrete advice based on rigorous research. The way they advertise and adapt their message to key actors, ranging from other experts, policymakers, journalists to ordinary citizens, is therefore a central component of their strategy.

This paper seeks to understand the ways in which specialized knowledge is disseminated and adapted in the discourse of American think tanks depending on the audience they are trying to convince. Although think tanks may use different modes of communication—from reports to social media – to advertise their research, the focus is on their blogs. Indeed, they constitute a widely accessible format and often draw on more extensive expert publications such as reports and policy briefs. It is argued that they are also a means for think tanks to carry out their programmatic aim in US society, since they are characterized by distinctive rhetorical strategies.

For think tanks, using blog posts to disseminate their expertise is still a fairly recent phenomenon. This implies that only a limited exploratory study could be conducted. The present study offers a comparative analysis of the rhetorical strategies used in two corpora composed of blog posts, on the one hand, and their corresponding expert reports, on the other hand. The objective is to shed light on a specialized community's multimodal discursive strategies by studying the way it adapts its rhetoric depending on its target audience. Considering that both blog posts and reports are designed according to their audience's needs, time availability and degree of expertise, they may be seen as serving different communicative functions in the discursive space of think tanks.

The paper has been structured as follows: section 2 offers a brief overview of think tanks' use of blogs to disseminate their research and the extent to which the latter medium conforms to the traditional features of blogs as identified by researchers. Section 3 introduces the two corpora as well as the methodology used to investigate the rhetorical strategies adopted by experts to fulfill their programmatic aim. Section 4 presents some preliminary results yielded by the study. Titles, hooks, rhetorical structure and hedging in both corpora are successively analyzed to highlight the different communicative functions of both genres, as expert voices meet, compete and adapt to the different readers they seek to influence. Section 5 offers concluding remarks.

2. An overview of think tanks' blogs

In a 2011 report to the *LSP Journal*, Rowley-Jolivet and Campagna (2011) present the Web 2.0 as a new medium to which specialized communities have had to adjust their discourse. This is particularly the case of blogs. Since they have multiplied over the last decades, new interpretative frameworks have had to be developed for their investigation. This section suggests that think tanks' blogs may constitute a genre on its own as they fulfill the milieu's specific objective.

2.1. Blogs and the notion of genre

The recent multiplication of blogs has led researchers to provide several typologies based on content (Herring *et al*. 2005), linguistic features (Grieve *et al*. 2011) or the rhetorical action they accomplish (Miller, Sheperd 2009). Miller and Sheperd (2009) thus distinguish the personal blog from the 'public affairs blog' depending on their social contexts and the corresponding rhetorical objectives of individual authors. Mauranen (2013) identifies two types of blogs, the 'personal' and the 'thematic'. The task of defining blogs as one genre is all the more difficult as they do not arise out of one discursive community whose members would be clearly identified. Their genre status has thus been questioned:

> Beyond the basic definition of blogging as the reverse-chronological posting of individually authored entries that include the capacity to provide hypertext links and often allow comment-based responses from readers, then, the term 'blog' now has little meaning unless a descriptive qualifier can be attached. (Bruns, Jacobs 2007, pp. 2-3)

It may therefore be more relevant "to talk about several blog genres rather than one" (Mauranen 2013), since they each emerge in specific communicative contexts to further individual or institutional objectives. In this respect, blogs appear to constitute more a medium with a set of technical possibilities rather than a full-fledged generic category defined by the recurrent social situation in which it occurs and by identified communicative purposes. Miller and Sheperd (2009, p. 273) identify the 'public-affairs blog' as a genre on its own because it arose in a post-09/11 era of distrust for mainstream media and participatory journalism. The technical capabilities of the medium met with a recurrent social need to provide alternative information, hence the emergence of a new blog genre.

Contrary to more traditional blogs created by one individual author for a community of anonymous readers and contributors, American think tanks' blogs are institutional products to which various experts in the organization contribute. Given their need to disseminate their expertise as widely as possible, think tanks first started to use blogs in the 2000s. The Heritage Foundation, often considered at the forefront of marketing techniques, created its own policy weblog in 2004. Older organizations soon followed.[1]

Though think tanks first used blogs mostly to circulate the op-eds and newspaper articles that their experts tried, sometimes unsuccessfully, to publish in US magazines, they have now expanded their uses of blogs, and today fully exploit the capabilities of the medium to disseminate their opinions. Blogs indeed "allow contributions such as reflective and controversial analysis which would rarely find their way on opinion pages" (Bahnisch 2007, p. 144), since newspapers often impose too many constraints

[1] One of the oldest think tanks in Washington DC, the Brookings Institution, has today more than seven different blogs, each of them dedicated to a specific theme dealt with by its experts.

on their contributors. They also allow for “making a sustained argument over time in a way that mainstream media rarely do” (Bahnisch 2007, p. 145) and thus enable think tanks to reach a wider audience, including ordinary citizens, who may never read their work otherwise. Because of their short and interactive format, they also respond to the evolving habits and needs of readers, all the more so as their target audiences, especially decision-makers, often lack time to sift through an ever-increasing quantity of information. As such, they fit into think tanks’ need for a segmented outreach strategy to convince different types of audiences (Selee 2013, p. 51).

Following Miller and Sheperd’s arguments (2009, p. 282-283), think tanks’ blogs thus epitomize the convergence between the milieu’s aim to influence and the technical affordances of a new medium, namely its user-friendliness and accessibility. To that extent, think tanks’ blogs may be envisaged as a genre on its own, emanating from the evolving communicative objectives of the producer community. The goal of this paper is to determine how the textual and digital features of the genre may serve the milieu’s objectives.

2.2. Characteristics of think tanks’ blogs

Although they respond to the specific needs of a specialized community, think tanks’ blogs are characterized by a set of features which for the most part conform to the characteristics identified by Bruns and Jacobs (2007, pp. 2-3) for the medium (see section 2.1.). Indeed, most of them are thematic; posts are organized in a reverse chronological order and classified according to the themes dealt with by experts. In the same way as the personalization of discourse has been identified as a feature of internet genres (Rowley-Jolivet, Campagna 2011, p. 48), think tanks’ blog posts are characterized by the use of the auctorial “I” and the names, contact details and portraits of the authors in the form of hypertext links. Surprisingly, the collaborative dimension (Bruns, Jacobs 2007, pp. 2-3; Rowley-Jolivet, Campagna 2011, p. 46) is not specific to the genre, for only one of the six think tanks’ blogs analyzed for this paper has a comment section.

What does seem specific to the genre however is the use of posts alongside longer publications. In addition to their initial commenting function similar to that of a newspaper opinion piece and which is actually quite common in blogs (Bondi, Seidenari 2015, p. 18), posts have increasingly been used by think tank experts to broadcast the results of their research. In a short period of time, ranging from a few days to a few months, an individual author may publish a report intended for expert readers and policymakers’ staffs and a blog post designed for a wider audience on the exact same topic with similar conclusions. Repackaging the research in different modes of communication is indeed part of the institutional strategy:

> The same research can well be the basis for outreach to members of Congress, business leaders, journalists and the public at large, but the way it is packaged is likely to change to suit the particular needs of each group and how they consume information. (Selee 2013, p. 51)

Authors can thus convey their message more efficiently and to a variety of readers, by providing concurrent media of different lengths or reading modes. Although the identity of blog readers is not clear, two types of readers have nonetheless been identified thanks to interviews[2] conducted with a number of think tank members. The first and most important group of readers consists of key members of the policy community in which they work; policymakers, journalists and other experts or opinion leaders are therefore the primary target. A second, larger and fuzzier category of readers is composed of well-educated and probably politically-aware citizens.

Because they address two different types of readers with varying degrees of expertise, blog posts are not so much considered by think tank members as a way to popularize their expertise, especially as they do not fulfill an informative or didactic purpose (Gotti 2003, p. 293). As evidenced by the absence of comment sections, the objective is not to interact with or answer the questions of readers. Indeed, there are very few instances of reformulation, explanation or de-terminologization processes typically associated with popularization (Mortureux 1985; Beacco, Moirand 1995; Gotti 2014) in blog posts.

Though they may consequently be seen as a mere summary of a more extensive piece of research in a more accessible format, blog posts are used by think tanks in a very specific way, with a view to serving their programmatic aim; in this respect, blog posts can be said to fulfill a distinctive communicative strategy. The following hypothesis is therefore explored: being designed according to the needs, time availability and degree of expertise of a heterogeneous community of readers, think tanks' blog posts are characterized by specific argumentative strategies. These rhetorical features may be considered as clues to the specific communicative function of blogs in the discursive space of think tanks.

The corpus and methodology adopted for the study are presented in the next section.

[2] I have conducted these ethnographic interviews in the context of my doctoral research on American think tanks' discourse between January 2016 and July 2017. More than 15 think tank members from various organizations and disciplinary backgrounds accepted to answer questions on the conditions of production of their discourse.

3. Corpus and methodology

3.1. Corpus

A small corpus was assembled to determine how think tank experts adapt their discourse to their target audiences. It is composed of six pairs of reports and blog posts written by six different authors in six American think tanks between 2014 and 2017. Though the corpus may seem small,[3] it was meant to serve the purpose of an exploratory analysis aiming to shed light on some of the multimodal discursive strategies used by think tanks: it also allows to account for the contexts in which each text was produced. The comparative analysis of both genres was expected to provide clues to and highlight blogs' distinctive communicative purpose.

The six pairs were selected to account for the variety of cases in which blog posts may be associated with longer reports. Different types of reports and orders of publication[4] for each pair were chosen – blog posts are published before, after or at the same time as reports. Various topics and types of think tanks were also selected, as shown in Table 1.

[3] The two corpora consist of 5, 255 and 45, 593 words respectively.

[4] Although the chronology may have seemed important at the beginning of the analysis, interviews with think tank members showed that most reports had often already been written when blog posts were published; the reason for some apparent delays lie in discrepancies in editing processes for the different types of text.

Blog	Type of report	Publication order	Topic	Think Tank	Type of think tank
Up Front	Policy Brief	Simultaneous (August 2014)[5]	Poverty Global Development	Brookings Institution	Academic[6] think tank Unidentified ideology Washington DC
Cato @ Liberty	Policy Analysis	Blog– Report (June – July 2017)	Flood Insurance	Cato Institute	Academic think tank Libertarian ideology Washington DC
Next New Deal: The Blog of the Roosevelt Institute	Report	Simultaneous (August 2017)	Environment Social Inequality	Roosevelt Institute	Advocacy tank Progressive ideology New York
The RAND blog	Report	Report – Blog (July – August 2017)	Education	RAND Corporation	Contract-research organization Unidentified ideology Santa Monica, California
The Mackinac Center's Blog	Report	Report-Blog (Dec 2016 – Feb 2017)	Cigarette taxes	Mackinac Center	Advocacy tank Conservative ideology Midland, Michigan
RFF blog	Report	Simultaneous (June 2017)	Car gas emissions	Resources for the Future	Academic think tank Unidentified ideology Washington DC

Table 1
Blog posts and reports selected for the corpus.

3.2. Methodology

Due to the small size of both sub-corpora and the resulting difficulty of obtaining meaningful quantitative results, a qualitative comparative approach was chosen to assess the various rhetorical strategies used by think tank experts to convince their different target audiences.

Titles and hooks were first examined, as they pertain to the way an author attracts his or her audience's attention and invites them to read on. Although hooks are mostly used by journalists, think tank members are acutely aware of the need for "an engaging hook that responds to the

[5] This first 2014 pair was selected as a first exploratory step based on an interview with its author in April 2016. As more blog posts and reports were selected for the corpus, no significant discrepancies were noted between the 2014 pair and the 2017 pairs, hence its presence in the final corpus.

[6] The different types of organizations identified in the table are based on Weaver's 1989 typology of think tanks depending on their institutional forms and objectives.

audience's need to know something about the issue at hand" (Selee 2013, p. 60) in their publications. Hooks therefore usually appear at the beginning of opinion pieces, reports or policy briefs and are particularly important in the case of blogs since the first lines of the posts come out directly on the blog homepage and are immediately visible to the reader. The ways in which these hooks, as well as titles, are tailored to the readers of reports and blogs were thus carefully compared.

Differences in the rhetorical structure of both genres were also analyzed through a study of rhetorical moves. Each move may be seen as serving a communicative intention and their distinctive combination testifies to the overall communicative purpose of the genre (Bhatia 1993, p. 23). They might thus constitute clues as to the differing communicative purposes of both reports and blogs. A secondary level of analysis was included with a comparison of the types of arguments used by authors in both genres.

Finally, attention was given to the discursive markers of authorial subjectivity through a study of hedging in the corpus. As think tank members aim to influence the political debate, their credibility as experts is particularly important to strengthen the validity of their claims, especially when writing for demanding readers such as other experts or policymakers. As Hyland (1994, p. 241) notes, "in persuasive writing, hedges are important means of both supporting the writer's position and building writer-reader relationships". The way authors present their research, convince readers of its soundness and interact with them both in blog posts and reports is therefore particularly crucial. Though the term "hedge" was first introduced by Lakoff (1972, p. 195) as a primarily semantic concept to designate words or phrases "whose job it is to make things fuzzier", a pragmatic conception of hedging was preferred. According to Fraser (2010, p. 201), hedging is indeed a fundamentally rhetorical strategy used to mitigate the strength or truth-validity of certain claims and, as such, may take many forms, including lexical items, text organisation or even prosody. Therefore, it is impossible to draw clear-cut lists of hedges in that "no linguistic items are inherently hedges but can acquire this quality depending on the communicative context or the co-text" (Clemen 1997, p. 6). For the purpose of this paper, a broader approach to hedging was adopted. Of course, traditional hedges such as modals, impersonal forms or rhetorical questions were examined but features pertaining to the way a text or a paragraph is structured were also included in the analysis.

The digital dimension of blog posts also needs to be accounted for. As opposed to reports which are posted as PDF files on think tanks' websites, blog posts constitute a full-fledged web genre "where the text, due to its media constraints, becomes an interaction medium, used actively to navigate the website" (Askehave, Nielsen 2005, p. 127). Therefore, elements associated with blogs' "navigating mode" such as hypertext links[7] or the digital location of the post on the website were taken into account in the analysis.

As specialized discourse is produced by disciplinary or professional communities and is thus constrained by their specific objectives, values and conventions, it is believed that these various elements can only be explained by taking into account their contexts of production, their authors' objectives and readers' expectations. Therefore, a context-driven approach, based on interviews conducted as part of previous research with various members of the think tank community (see endnote 1), was adopted to further the analysis.

The following section is devoted to the results yielded by the comparative study of rhetorical strategies in both sub-corpora.

4. Results

4.1. Titles and hooks: attracting readers' attention

To shed light on the way think tanks adapt their programmatic discourse to their target audiences, readers' expectations and reading modes first need to be accounted for. Compared to readers of extensive reports who look for concrete solutions to a policy problem, most blog readers may only be interested in reading a short piece of information or analysis. Blog posts are accessible to all whereas reports are distributed to key stakeholders. Though both genres seek to be enticing to readers, these differences make it even more important for the information in blogs to be presented in such a way that it will catch the reader's eye almost immediately, hence the need for relevant titles and hooks.

4.1.1. Titles

According to interviews with members and editors of think tanks, an efficient title either in blogs or policy papers, just as for newspaper articles, is one that will not only summarize the content of the author's message clearly but might also slightly overplay it, thus arousing readers' curiosity and, in the case of blogs, inviting them to click on the post. However, they differ in terms of the specificity of the information they provide, which reflects the differing needs of their target audiences. Table 2 presents the titles in both sub-corpora.

[7] In italics in the examples developed in the next sections.

Think Tank	Report	Blog Post
Brookings Institution	How Poor are America's Poorest? U.S. $2 a Day Poverty in a Global Context	U.S. $2 a Day Poverty in a Global Context: Five Questions Answered
Cato Institute	Reforming the National Flood Insurance Program Toward Private Flood Insurance	Toward Private Flood Insurance
Roosevelt Institute	Boiling Points: The Inextricable Links Between Inequality and Climate Change	Inequality is Toxic: How Economic and Social Disparities Drive Climate Change
RAND Corporation	Observations and Guidance on Implementing Personalized Learning	What Emerging Research Says about the Promise of Personalized Learning
Mackinac Center	Cigarette Taxes and Smuggling: a 2016 Update	Cigarette Taxes and Smuggling: an update through 2015 *Original title: Do cigarette tax hikes make people kick the habit? Not necessarily.*
Resources For the Future	How Much Do Consumers Value Fuel Economy and Performance? Evidence from Technology Adoption	Do Consumers Benefit from Automobile Fuel Economy and Greenhouse Gas Standards?

Table 2
Comparison of titles in reports and blog posts.

Contrary to what might have been expected from shorter pieces, blog posts' titles are not necessarily catchier than those of reports. Nevertheless, while reports' titles provide general information on their content, authors' analyses or opinions in blog posts tend to be directly put to the fore, as may be seen with the play on the word "toxic" in the Roosevelt Institute's blog post, the transition from an open to a yes/no question in Resources For the Future (RFF)'s pair or the preposition "toward" in the Cato Institute's. The subtitle of the Cato report is here promoted as a title in the blog post even though the latter was published beforehand to coincide with the holding of a House Committee on flood insurance. By advocating his or her take on a timely issue from the title, the author can catch the eye of an informed audience who may have heard about the debate – in this case probably policymakers and journalists.

Blog posts' titles may also use words echoing current and possibly controversial issues. This is the case of the RFF post on the potential benefits of fuel economy and greenhouse gas standards for cars; actually, its release coincided with the US withdrawal from the Paris agreement on climate change. The key words "automobile" and "greenhouse gas" link the blog post to the current context and make it all the more clickable. Although the report was released at the same time, its title does not align with current public concerns but with academic matters.

Blogs also contribute to maintain a community of regular readers (Mauranen 2013) and titles can be tailored accordingly. This is particularly striking in the case of the Michigander Mackinac Center's post which was initially published in a Washington DC political magazine *The Hill* under a catchier and more news-worthy title to promote the results of an annual study on cigarette taxes. When reproduced on the Mackinac Center's local blog, the new title "An update through 2015" might have been deemed more appropriate to draw the attention of regular readers who might have read the previous reports.

4.1.2. Hooks

Just as titles in blog posts seem designed to attract their intended audience's attention by using key words or putting forward authors' opinions, hooks are tailored to respond to public concerns. Indeed, although most think tanks' publications start with this journalistic technique, "each may have slightly different hooks depending on the intended audience" (Selee 2013, p.60). Those are presented in Table 3.

Think Tank	Report	Blog Post
Brookings Institution	Shocking findings of economists Schaeffer and Edin's recent study	*Editor's note*: recent studies by Piketty and others
Cato Institute	A problematic program on flood insurance set to expire at the end of September	A bill on flood insurance examined the following day by a House Committee
Roosevelt Institute	The discrepancy between the absence of conversation on climate change as opposed to the one on inequality	The People's Climate March in April
RAND Corporation	Little research done on a recent phenomenon (personalized learning)	The discrepancy between the author's 2015 research and his newer research on personalized learning
Mackinac Center	Unforeseen consequences in policy making	*Editor's note*: A piece published in a DC newspaper The discrepancy between common belief on how taxes may change behavior and results of recent research
Resources For the Future	The objectives of the Environmental Protection Agency (EPA) as a background for the study	The approaching deadline for EPA to finalize fuel economy and greenhouse gas emissions standards

Table 3
Comparison of hooks in reports and blog posts.

In reports, hooks tend to rest on the identification of a problem that has been on the target audience's mind. Usually, this audience consists of other experts or stakeholders looking for an applicable solution to a policy issue. Hooks in reports may therefore point out gaps (RAND Corporation, Roosevelt Institute) or surprising findings (Brookings Institution) in applied policy research, or else remind readers of an upcoming or on-going policy discussion (Resources For the Future; Cato Institute). They thus refer to shared knowledge: most experts in global development have for instance heard about the Edin and Schaeffer study (Brookings Institution) or are aware of a policy issue in their field of expertise.

While blog posts' structure rests on the same rhetorical strategy (i.e. the identification of a policy problem) as will be shown in the next sub-section, their hooks are linked to a wider audience's concerns. Those may be news hooks such as upcoming deadlines (Cato Institute; Resources For the Future), recent events (Roosevelt Institute) or people most readers of the blog have heard about (Brookings Institution). The role of editor's notes should also be considered as they are part of the think tank's strategy to draw attention to their experts' work. Well-known economist Thomas Piketty is mentioned right at the beginning of the Brookings post even though he is actually quoted only once in the conclusion of the corresponding report. An editor's note is also added to the Mackinac blog post to remind its readers that the piece has been published in a Washington DC magazine, thereby valorizing the organization's reputation at the national level and indirectly attracting local readers.

By tailoring the titles and first lines of their blog posts to a wide audience's concerns, think tank experts hope to attract readers' attention and invite them to read their proposals. While reports seem to be rooted in long-term discussions involving mostly experts and stakeholders, titles and hooks contribute to set blogs in current, day-to-day policy debates likely to resonate in many informed readers' minds. Blogs thus fulfill their primary objective, that of disseminating the contents of the think tank's work. A secondary and more distinctive communicative intention for the genre may be identified through a comparison of argumentative strategies in both reports and blogs.

4.2. Rhetorical moves: highlighting results and analysis

While each rhetorical move may serve a communicative intention in a given text, it is their unique combination which allows for the identification of the text's communicative purpose and consequently its affiliation to a specific genre. The report genre is to that extent easily identified, since it is characterized by a set of recurrent rhetorical moves designed to meet the expectations of demanding readers looking for concrete solutions to a policy issue. On the contrary, apart from the aforementioned hooks, blog posts have no fixed structure and generally follow reports' lines of argumentation. Contrasting their argumentative content with that of their corresponding reports might nonetheless shed light on their distinctive communicative purpose.

Argumentation in reports usually rests on the identification of a policy issue where a discrepancy between what is and what should be is pointed out. Not only can the author thus justify the relevance of his or her research and therefore of the present report, but he or she can also propose a series of concrete recommendations for policymakers to consider. In the case of the Brookings Institution report, the aim of the author is to show how current methods to measure poverty at a global level, as those used by organizations such as the World Bank, are not relevant because they cannot be applied meaningfully across diverse settings including the United States. Having identified the limits of such measures, the report then exhibits the following moves: a detailed analysis of current methods to measure poverty first in the US, then in the world, is offered, followed by a presentation of potential alternative methods, and recommendations for international organizations to consider. Regarding the blog post, its argumentative structure rests on the same policy issue but is organized following five rhetorical questions a wide audience may have on poverty in the US and the world. Nevertheless, each answer corresponds to summaries or to the final analysis of each part of the report. The research background for the study, the methodology that may have been used and the recommendations which are discussed extensively in the report are almost absent in the blog post as shown in Figure 1.

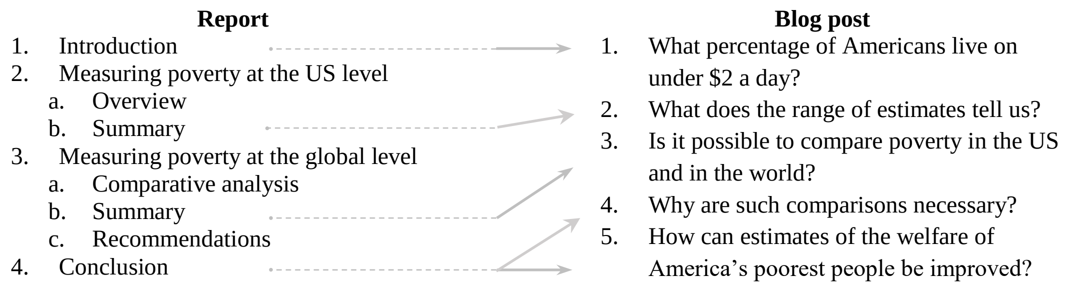


Figure 1
Comparison between rhetorical moves[8] in the Brookings Institution's report and blog post.

The same phenomenon may be observed for the Mackinac Center's report and blog post on the unintended consequences of cigarette taxes on smuggling in the United States. Whereas the background for the study, its method, results, and policy recommendations are successively introduced in the report, only a few elements of methodology and results remain in the post. In the same way, all anecdotes on the negative impacts of cigarette taxes on smuggling used in the report to provide talking points for policymakers are surprisingly absent in the post.

[8] The moves identified for the Brookings blog post correspond to the subtitles used by the author. The subtitles used in the report do not correspond exactly to its rhetorical moves and therefore have been adapted in Figure 1.

Of course, such transition from reports to blogs may simply be due to the format of posts which only allow for about 800 words. However, the emphasis on personal analysis and results may also be interpreted as a way for authors, not only to summarize the content of their research but also to strongly voice their claims, justify their relevance and draw the attention of busy readers who often have to sift through hundreds of similar posts and articles every day. The way these claims are brought about provides further clues to blogs' specific communicative intent in the discursive space of think tanks.

Interestingly, arguments in blog posts seem to have been specifically selected for their rhetorical strength in the eyes of a wide audience. Although the Roosevelt Institute report studies the links between climate change and inequality, its corresponding post focuses on its most striking finding, that is that inequality is a cause of climate change. Besides, only examples drawn from studies on the United States or by famous economists such as James Boyce remain in the post. These are indeed the most telling for a community of mostly American well-informed left-leaning readers.

Readers' political beliefs may actually be a key criterion in the way advocacy think tank experts, such as the Roosevelt Institute's or the Cato Institute's, adapt their message in blog posts. This can be seen in the following example comparing the libertarian Cato Institute's report with the blog post:

(1)	**Report**	**Blog post**
	The Biggert-Waters backlash demonstrates the applicability of core public choice insights about public policy in modern democratic states to the NFIP. Public choice would say that policy often concentrates benefits on a small, vocal interest group while spreading corresponding costs across society as a whole. The combination of concentrated benefits and diffused costs makes rolling back policy much more difficult than enacting it […]. (Cato Institute, 2017)	For instance, in 2012 Congress passed the *Biggert-Waters Flood Insurance Reform Act*, which required the NFIP to end subsidies and to begin including a catastrophe loading surcharge. However, due to interest group pressure, Congress *reversed itself* just two years later, halting some reforms and getting rid of others outright. The quick backtrack was a classic example of government failing to act in the public interest due to *concentrated benefits and diffused costs*. (Cato Institute, 2017)

In the report, as suggested by the modal "would", public choice theory is presented as one among several potential interpretations of Congress's step back on the bill – even though it is the author's favourite interpretation. In the blog post, on the other hand, the argument is framed as another example of the failure of government intervention – a recurring theme in libertarian ideology and a view which is probably shared by readers of the Cato Institute blog. While reports seek to convince demanding readers and therefore need to be built on objective evidence, blog posts may be aimed at regular readers of the think tank's work who often share their political opinion. Presented with arguments they are familiar with or feel strongly about, blog readers are more likely to react and want to know more.

Paradoxically, the removal of details such as recommendations in blog posts may prompt certain readers of the blog such as decision-makers or opinion leaders to want to know more and read the full report. Hypertext links to the report on which the post is based or to the profile of the author, his or her other publications and contact information enable interested decision-makers to read the report or, more probably, ask for a private briefing on the issue at hand.

As their arguments are aligned with the needs and concerns of a wide audience, blog posts seem to be characterized by different rhetorical strategies from those of reports. The emphasis on personal analysis and results with a view to catching the attention of a variety of readers may indeed testify to blogs' specific rhetorical function for think tanks – that of a stage from which think tank experts may position themselves in a crowded marketplace of ideas.

4.3. Hedging: finding a voice in a crowded political arena

With an estimated number of 1, 835 think tanks in the United States seeking to influence public policy, about a quarter of which are located in Washington DC (Think Tanks and Civil Societies Program 2018), think tank experts have to compete with a multitude of different voices. New media have actually contributed to the phenomenon:

> The advent of social media including blogs, Facebook, and Twitter, has vastly increased the opportunities for outreach, but these new technologies have also created a more crowded marketplace for ideas that organizations have to contend with. (Selee 2013, p. 13)

The way experts put forward and frame their claims, but also position themselves and hedge their arguments – especially in blog posts where such marketplace is bigger and more heterogeneous – is therefore crucial. Although the corpus is too small to draw any definitive conclusion about hedging in both genres on a quantitative level, it is believed that a qualitative comparison of the way think tank experts hedge their claims may provide further clues as to their stakes.

While in reports readers look for rigorous analysis based on evidence to decide on a course of action, blog readers have no need for action and may lack expertise. Authors' scientific credibility is therefore less at stake in blogs than in reports. As a consequence, think tank experts seem to use fewer traditional hedges in blogs when it comes to presenting their personal analysis as the following examples illustrate:

(2)

Report	**Blog post**
Yet the results of this study suggest that personalized learning has positive effects. (RAND Corporation, 2017)	Personalized learning continues to look promising because positive effects were evident in the 2017 study despite these hurdles. (RAND Corporation, 2017)

(3)

Report	**Blog post**
From the heat islands of U.S. cities to rural farming communities, the poorest groups in America, many of them communities of color**,** will likely experience the worst effects of climate change but have the least ability to cope with and adapt to it. (Roosevelt Institute, 2017)	In inner cities, Native reservations, and rural farming communities, the poorest groups in America, many of color**,** are experiencing the worst effects of climate change but have the least ability to cope with and adapt to it. (Roosevelt Institute, 2017)

Not only do think tank experts tend to be more assertive but they are also more likely to simplify their arguments:

(4)

Report	**Blog post**
Wealthier households benefit disproportionately from the reduced average cost of flood insurance brought about by government intervention. Of course, not all NFIP-insured properties are high value, but insured homes are on average more valuable than noninsured homes. (Cato Institute, 2017)	Because the average home in the NFIP is much more valuable than an average American home, the program is regressive on the whole. (Cato Institute, 2017)

Although authors use fewer traditional hedges in blogs posts when making their claims, a pragmatic approach to hedging actually shows that this does not mean that they do not protect their own faces as rigorous experts. As evidenced in examples (1) and (5), hypertext links in italics, when referring to external research studies or government websites, may fulfill a hedging function. When interviewed on her use of links in blog posts, one member of the Cato Institute noted:

> [T]he reason I do that is obviously because it provides additional evidence to people who are questioning or may not take what I'm saying at face value. If they want to investigate a claim that I'm making in the article, then there is an ability to do that. (Interview TT-11, Cato Institute, 30th June 2017)

In her view, hypertext links support her credibility as an expert in her own field, which by extension allows her to take a stance in the marketplace of ideas.

Considering that the political arena is wider and more heterogeneous in the case of blogs, authors may actually be driven to hedge against misguided interpretation and boost their claims. For example, RFF research shows that consumers do not benefit from tighter standards on fuel economy since the benefits of fuel economy are wasted in the loss of vehicle performance. In their blog post, the authors insist however on the fact that such results should not be misinterpreted by climate change skeptics as tighter standards may still have an impact on the environment. Such caveat is not put forward as much in the report and is in fact part of a larger section dedicated to the limitations of the study:

(5)

Report	**Blog post**
This conclusion is subject to the caveats we discuss in Section 5.2, and we note that standards may increase social welfare after accounting for the energy security and climate benefits. (Resources For the Future, 2017)	We emphasize that these results do not mean that the tighter standards have harmed the public. […] Given the possibility that the energy security benefits of lower oil consumption *may be lower now than they used to be*, our results suggest that the chief benefits of tighter standards may be to reduce US greenhouse gas emissions and help lay a foundation for international efforts to reduce the costs of climate change. (Resources For the Future, 2017)

What may explain such rhetorical precautions in the post is the fact that it was published at a time when President Trump was considering withdrawing from the Paris agreement on climate change. While other experts possess the necessary background to understand what the results of a study in their field may or may not suggest, blog readers are not necessarily as informed on the subject and have to make up their own minds on an extremely controversial issue based on the limited information they have.

Think tank experts contend with more numerous and consequently more divergent voices in blog posts. They face the risk of being criticized on scientific but also – and mostly – on political grounds, hence the need to explain what the study is exactly about or, more generally to put forward their personal analysis from the title and boost their claims. In example (2), a RAND expert discusses his research on personalized learning, the results of which were actually not as conclusive as expected. This may explain why he felt the need to insist on its relevance in the post through boosters such as "evident" which allows him to mark his certainty in what is being said or a more explicit title (see Table 1). This may have been all the more necessary as it was also published on a Brookings Institution's blog dedicated to education and therefore directly targeted towards the education community.

Since they are published within a short time-frame determined by news and current events, blogs are particularly suited to experts' needs to position themselves. When questioned about her manager ([TT-4])'s uses of blogs, a research assistant at the Brookings Institution explains:

> Blog posts are also opportunistic as some things will come up that apply to the things we're doing…[TT-4] is positioning herself as an expert in her field and so if something really big happens, she wants to be…or should be the source for quality information on that topic. She'll write a blog post so that's responding to the big things that are happening in this area. (Interview TT-3, Brookings Institution, 26th April 2016)

While reports allow think tank members to position themselves in an expert community where scientific rigor and objectivity are paramount, blogs appear as a means to take a stance in the political arena. Think tank experts may hedge that stance either by putting forward the results of their analyses and/or defending their strength in the face of potential criticism. Political relevance is thus a distinctive stake for blogs.

5. Conclusion

The purpose of this paper was to shed light on some of American think tanks' multimodal strategies to disseminate their expertise in order to fulfill their programmatic objective. Resulting from a convergence between the technical capabilities of an accessible medium and a specialized community's need for outreach, think tanks' blogs have been identified as one of the milieu's preferred genres to distribute its research to a wide, though heterogeneous, community of readers. Yet, think tanks' blogs do not constitute a mere summary of expert publications but rather fulfill think tanks' goal to influence the decision-making process as their format, timeliness and content are tailored to the needs of their target audience.

The comparative analysis of rhetorical features in both reports and blogs has highlighted a distinctive and "complexly layered set" of communicative purposes for the genre (Askehave, Swales 2001, p. 199).

Journalistic techniques such as titles and hooks designed to address readers' concerns allow authors to draw attention to their work while justifying its relevance in day-to-day political debates. Argumentative choices are framed to resonate with a well-informed and, in some cases, politically-aware audience and therefore reflect the genre's specific programmatic aim. The emphasis on personal analysis, as highlighted through a pragmatic approach to hedging in both corpora, may be seen as a way for authors to take a stance in a crowded political arena where they compete and/or contend with similar actors. Blogs may thus constitute a means for think tank experts to position themselves, either by advocating for new ideas or reacting to potential criticisms. More generally, they represent a window into the organization's work, thereby contributing to the institutional image as having an impact on the policy process.

Although this would require further research with a larger corpus, these overlapping communicative purposes set blogs as a distinctive genre in think tanks' outreach strategy to influence the decision-making process and testify to the specialized nature of the genre. Finally, the importance of a context-driven approach to specialized discourse should be underlined, since it has allowed to shed light on new practices as communities adapt to evolving social contexts.

Bionote: Mathilde Gaillard is a fourth-year PhD student at the Centre de Linguistique en Sorbonne at Paris Sorbonne University under the supervision of Professor Catherine Resche. Her doctoral research focuses on the discourse of a particular professional community—American think tanks—through a two-pronged analysis of their discursive genres and of the cultural conditions in which these are produced. She is particularly interested in the impact of think tanks' programmatic aim on their rhetorical choices in discourse. She is also a lecturer at Paris Nanterre University where she teaches English for students in applied foreign languages.

Author's address: mathilde.gaillard@ens-cachan.fr

References

Askehave I. and Swales J. 2001, *Genre identification and communicative purpose: a problem and a possible solution*, in "Applied Linguistics" 22 [1], pp. 195-212.

Askehave I. and Nielsen A. 2005, *Digital genres: a challenge to traditional genre theory*, in "Information Technology & People" 18 [2], pp. 120-141.

Bahnisch M. 2007, *The political uses of blogs*, in Bruns A. and Jacobs J. (eds.), *The uses of blogs*, Peter Lang, New York, pp. 137-149.

Beacco J-C. and Moirand S. 1995, *Autour des discours de transmission des connaissances*, in "Langages" 117, pp. 32-53.

Bhatia V.K. 1993, *Analysing Genre: Language Use in Professional Settings*, Longman, London.

Bondi M. and Seidenari C. 2015, *And now I'm finally of the mind to say i hope the whole ship goes down...: Markers of subjectivity and evaluative phraseology in blogs*, in Mukherjee J. and Huber M. (eds), *Corpus Linguistics and Variation in English: Theory and Description*, Amsterdam-New York, Rodopi, pp. 17-27.

Bruns A. and Jacobs J. (eds.) 2007, *The uses of blogs*, Peter Lang, New York.

Fraser B. 2010, *Pragmatic competence: The case of hedging*, in Gunther K., Mihatsch W. and Schneider S. (eds.), *New Approaches to Hedging*, Emerald, Bingley, pp. 15-34.

Gotti M. 2003, *Specialized discourse: linguistic features and changing conventions*, Peter Lang, Bern.

Gotti M. 2014, *Reformulation and recontextualization in popularization discourse*, in "Ibérica" 27, pp. 15-34.

Grieve J., Biber D., Friginal E. and Nekrasova T. 2011, *Variation among blog text types: A multi-dimensional analysis*, in Mehler A., Sharoff S. and Santini M. (eds.), *Genres on the Web. Computational Models and Empirical Studies*, Springer-Verlag, New York, pp. 303-322.

Herring S., Scheidt L.A., Wright E. and Bonus S. 2005, *Weblogs as a bridging genre*, in "Information Technology & People" 18 [2], pp. 142-171.

Hyland K. 1994, *Hedging in Academic Writing and EAP Textbooks*, in "English for Specific Purposes" 13 [3], pp. 239-256.

Lakoff G. 1972, *Hedges: A study in meaning criteria and the logic of fuzzy concepts*, in "Papers from the Eighth Regional Meeting of the Chicago Linguistic Society", pp. 183-228.

Mauranen A. 2013, *Why take an interest in the research blogging?*, in "The European English Messenger" 22 [1], 53-58.

Mortureux F. 1985, *Linguistique et vulgarisation scientifique*, in "Social Science Information" 24 [4], pp. 825-845.

Rowley-Jolivet E. and Campagna S. 2011, *From print to web 2.0: The changing face of discourse for special purposes*, in "LSP Journal" 2 [2], pp. 44-51.

Selee A. 2013, *What Should Think Tanks Do? A Strategic Guide to Policy Impact*, Stanford University Press, Stanford.

Think Tanks and Civil Societies Program. 2018, *2017 Global Go To Think Tank Index Report*, Philadelphia, The Lauder Institute, University of Pennsylvania.

Weaver R.K. 1989, *The Changing World of Think Tanks*, in "PS: Political Science and Politics" 22 [3], pp. 563-578.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 359-382
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p359
http://siba-ese.unisalento.it, © 2019 Università del Salento

GOOGLE TALKS AS A NEW KNOWLEDGE DISSEMINATION GENRE

ELISA MATTIELLO
UNIVERSITÀ DI PISA

Abstract – This study explores the new knowledge dissemination (KD) online genre of Google Talks, in both qualitative and quantitative terms. In particular, the study combines two complementary strands of linguistic investigation – discourse analysis and corpus analysis – to inspect and describe the features that characterise Google Talks as popularisation discourse, as compared to both traditional and new web-based genres. The qualitative analysis of three case studies belonging to the fields of economics, political science, and medicine shows both a continuity between Google Talks and other forms of popularisation, such as TED Talks, and a departure from more traditional genres in academic and institutional settings addressed at non-experts (academic lectures) or colleagues (conference presentations). A quantitative corpus-based analysis of evaluative adjectives shows that Google speakers frequently use aesthetic and emotion adjectives to encourage audience participation and create intimacy and proximity with hearers. In general, Google Talks imposes not only a simplification but also a reformulation and recontextualisation of specialised knowledge in a more interactive and dynamic web-based context.

Keywords: Google Talks; knowledge dissemination (KD); popularisation; evaluative adjectives; web genre.

1. Introduction

Nowadays the increasing importance of knowledge dissemination (KD) has led to the emergence of a wide array of genres – from newspaper or journal articles to more recent web-mediated genres such as TED Talks, science blogs, and Social Networking Sites (e.g. Facebook, Twitter). New genres on the World Wide Web, with their growing level of participation and interaction, cater to the necessity of reaching a wide global audience and of making specialised knowledge accessible also to non-experts (or experts in other fields) (Caliendo 2012; Compagnone 2014; Garzone 2012; Luzón 2013; Mauranen 2013; Myers 2010).

This study focuses on Google Talks, i.e. a new online genre which consists of a series of video presentations by invited speakers given at various Google offices throughout the world and available on the website *Talks at Google. Where Great Minds Meet* (https://talksat.withgoogle.com/). On this website, Google speakers are manifold, including expert CEOs and physicians, but also well-known celebrities (e.g. singers, actors, athletes), or famous politicians. The talks generally last about fifty minutes, but their length can vary, especially depending on the final discussion session, which follows the initial monologic part. No transcripts of the talks are available on the website.

The aim of this study is to highlight the features of Google Talks and the linguistic strategies adopted (and adapted) in relation to stakeholders from different cultural backgrounds, either experts or non-/semi-experts. In particular, the study explores three case studies of Google Talks belonging to different domains, namely business/economics, political science, and medicine. For the corpus-based analysis, the emphasis is on evaluative (subjective) adjectives and their contribution to reinforcing meaning and creating "proximity" (Hyland 2010) with the audience.

In the study, the following research questions will be addressed:

1. Which features characterise Google Talks as a new genre of popularisation discourse and how does it differ from other KD genres?
2. What are the incidence and role of subjective adjectives in the processes of (a) speakers' expression of epistemic stance and (b) hearers' engagement?

In Google Talks, speakers constantly operate within and across generic boundaries creating a new hybrid form of text. In particular, Google Talks provides evidence of "Genre-Mixing" (see Bhatia 1995, 2012), in that it mixes features of both established genres, such as university lectures or research talks, and emerging popularisation genres, such as TED Talks.

2. Theoretical framework and other oral genres

2.1. The theoretical framework

For the analysis of Google Talks, this study combines two complementary approaches: i.e. 1) discourse and genre analysis, with special attention to new digital genres of popularisation discourse, and 2) corpus-based analysis.

For the qualitative analysis, the study adopts a discourse and genre analysis approach (Bhatia 1993, 2004, 2012; Fairclough 2003; Swales 1990, 2004). In particular, for Discourse Analysis, it draws on Fairclough's (2003) social, discourse and textual analysis, with special focus on the concept of "intertextuality", that is, reliance on prior texts and text types (de

Beaugrande, Dressler 1981). It investigates the production of meaning using linguistic features, rhetorical strategies, and other non-verbal semiotic resources, with some attention paid to the communicative purposes that the new genre of Google Talks tends to serve and the professional context it is situated in.

As for Genre Analysis, it draws on Swales' (1990, 2004) and Bhatia's (1993) development of genre theory to analyse academic and professional genres, with specific focus on the identification of the qualities characterising Google Talks as a hybrid or mixed genre that allows speakers to give expression to their private intentions (Bhatia 1995). A propos, Bhatia (2012, p. 24) introduces the concept of "interdiscursivity", which can be viewed as a function of "appropriation of generic resources" across three kinds of contextual and/or text-external resources: i.e. genres, professional practices, and professional cultures. Thus, while intertextuality operates within the textual space and concerns appropriations across text-internal resources, interdiscursivity concerns appropriations across text-external semiotic resources, such as genres, professional, institutional, and disciplinary practices. These appropriations simultaneously operate at all levels of discourse to realise the intended meaning, and have been widely used in the "recontextualisation" or "reformulation" (Calsamiglia, van Dijk 2004; Gotti 2014) of existing discourses and genres into novel or hybrid forms. The diagram representing Bhatia's (2012) "Interdiscursivity in Genre Theory" is provided in Figure 1.

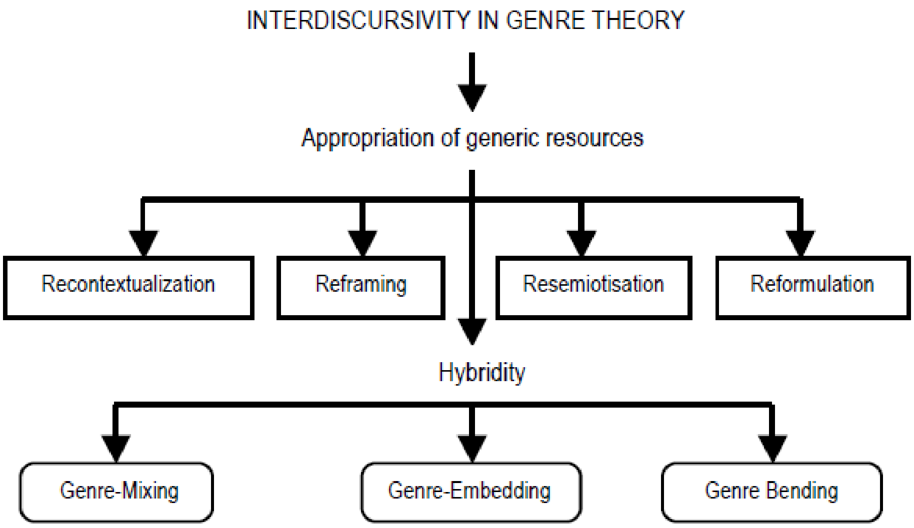


Figure 1
Interdiscursivity in Genre Theory (Bhatia 2012, p. 25).

In particular, the hybridity of Google Talks results from "Genre-Mixing", in that this new digital genre mixes features of various others, not only widely studied traditional ones, such as university lectures (Artiga León 2006;

Bamford 2009; Caliendo, Compagnone 2014; Crawford Camiciottoli 2008, 2015, 2016; Fortanet 2004; Walsh 2004), but also genres of popularisation discourse, such as TED Talks and other web-mediated genres of popularisation (Caliendo 2012; Compagnone 2014; Garzone 2006, 2012; Mauranen 2013; Myers 2010; Scotto di Carlo 2013, 2015). The next section is dedicated to the description of the main features of university lectures, conference presentations, and TED Talks as comparable KD genres.

2.2. Other oral genres

In the last few decades, scholars have focused their interest on the way experts disseminate specialised knowledge both to their peers and to non- or semi-experts. Among widely explored oral genres, traditional genres, such as university lectures and conference presentations or research talks, have been studied along with emerging popularisation genres, such as TED Talks or blogs, as a way to identify their main features and functions.

The genre of academic lectures, for instance, has attracted the attention of several scholars, whose focus was especially on the use of first and second person pronouns (Fortanet 2004; Walsh 2004), phraseology (Artiga León 2006), evaluation (Bamford 2009), and epistemic lexical verbs used to promote an interactional approach (Caliendo, Compagnone 2014). Studies dedicated to lecture discourse have highlighted numerous linguistic features used by lecturers to explain disciplinary concepts and enhance the novice's understanding. Among the latter are interactional devices, i.e. comprehension checks, questions, imperatives, as well as language features linked to informality, including idioms and puns.

Recently, Crawford Camiciottoli (2015, 2016) has also adopted a multimodal approach to investigate the interplay of verbal and non-verbal strategies in OpenCourseWare humanities lectures. Remarkably, her findings show how verbal (e.g. questions, humour) and non-verbal features (e.g. prosodic stress, gaze direction, and hand/arm gesturing) can work synergistically in university lectures to improve comprehension and promote a learning-friendly classroom atmosphere.

Another type of oral genre which has attracted the attention of scholars is the conference presentation. Conference presentations are planned speech events, which are organised and prearranged to comply with the time slot provided. According to Jurado (2017, p. 46), conference presentations can be classified as "an academic genre" since it is mainly used by discourse communities within academia in order to present a scientific novelty, give visibility to research, and reinforce social cohesion within the discourse community (Rowley-Jolivet 1999, p. 179). Carter-Thomas and Rowley-Jolivet (2003) also notice that conference presentations are mainly informative, displaying high-density informational content, but at the same time persuasive, thus implying the use of rhetorical devices.

An issue of particular relevance raised in the literature on conference presentations is the immediacy in time and place in relation to an expert audience with whom a great amount of shared knowledge can be expected. For instance, Hood and Forey (2005) and Rowley-Jolivet and Carter-Thomas (2005) maintain that the presence of an audience and the consequent need for real-time processing strongly influence the way a presentation is designed (e.g. less dense, simpler, with a considerable amount of signposting and inclusive use of pronouns).

Hood and Forey (2005) also notice how the use of verbal features, such as evaluation, and non-verbal features, such as inclusive gestures, can contribute to the presentation design. Moreover, as noticed by Bucher and Niemann (2012) and Hertz (2015), recent research talks are commonly PowerPoint-supported by slideshows or similar visuals, allowing speakers to be better understood and arousing higher interest among participants.

There are also some interesting studies that deal with specific sections of conference presentations. For instance, the discussion (or Q&A) session has been dealt with by Wulff, Swales and Keller (2009), who consider it as part of the genre, but still notice some important differences, namely, while the presentation is more prepared and closer to written language, the discussion session is more conversational and closer to spoken language.

Another particular type of research dissemination genre which has recently gained the attention of scholars is TED Talks (Caliendo, Compagnone 2014; Compagnone 2014; Masi 2016 *inter alia*). Caliendo and Compagnone (2014, p. 105) define this genre as "a series of short popularizing talks (of approximately twenty minutes), addressing a mass audience and delivered by top-level experts in a wide variety of domains". They consider TED Talks as an on-line genre which provides experts with the chance to disseminate knowledge outside their disciplinary communities, both to a physically present audience, and to the web-users at home.

The literature on popularisation has explored TED Talks from different angles, highlighting their characteristics. Scotto di Carlo (2013, 2015), for instance, has investigated TED speakers' use of humour and subjective adjectives to establish a connection with the audience, while Mattiello (2017) has analysed TEDsters' use of simple vocabulary and informal register as strategies of science popularisation. In the latter study, the qualitative analysis also stresses the importance of humour and narration for increasing speaker-hearer empathy.

Among web-mediated genres, TED Talks has been described as a "new hybrid genre" (Caliendo 2012, p. 101) that is gaining more and more interest also in the area of screen-mediated communication and multimodal literacy. Interestingly, Caliendo (2012) and Compagnone (2014) have acknowledged the multimodal nature of this genre, noting that it mixes different semiotic modes (i.e. spoken, written, video, and audio), and Masi (2016) has recently highlighted the important contribution of non-verbal devices, principally deictic and metaphoric gestures, to the meaning of TED Talks in the field of economics.

Since Google Talks seems to share some features with each of these oral genres, it is the purpose of the analysis conducted in this paper to identify which characteristics are shared and which instead qualify it as a new autonomous genre of knowledge dissemination. The following section describes the corpus and methodology.

3. Corpus and methodology

3.1. The corpus

The corpus selected for the analysis includes three Google Talks available on the *Talks at Google* website.

On *Talks at Google*, areas or domains can be selected either by category (e.g. 'Economics', 'Health & Wellbeing'), or, alphabetically, by topic (e.g. 'Business' under B, or 'Capitalism' under C). Videos are chronologically ordered from the most to the least recent one. For this study, I have selected recent talks belonging to three different domains:

- Business/Economics: "Superbosses: How Exceptional Leaders Master the Flow of Talent" [10,994 words], by Sydney Finkelstein, delivered on 14th July 2016, duration 50:50;
- Politics: "Naked Diplomacy" [9,328 words], by Thomas S.F. Fletcher, delivered on 7th June 2016, duration 46:32;
- Medicine/Health: "Anatomy of a Breakthrough in Targeted Cancer Treatments" [15,034 words], by Brian J. Druker, delivered on 1st June 2015, duration 01:20:04.

The overall corpus totals 35,356 tokens and is drawn from a larger corpus that is being collected for a national research programme.[1]

3.2. The method

In order to highlight the main features characterising Google Talks, the three talks selected were first investigated from a qualitative perspective. The

[1] This research has been financed by the Italian Ministry for the University (PRIN 2015 no. 2015TJ8ZAS). Other fields of interest for the national PRIN programme are Law, Technology, and Tourism.

multimodal analysis focused on both verbal strategies, such as informal register, figurative language, and narratives, and non-verbal strategies, such as hand/arm gestures, head/body movements, and gaze direction (Baldry, Thibault 2006; Kress, van Leeuwen 1996). For the investigation of gestures, reference was made to McNeill's (1992) classification, covering 1) beats, i.e. repeated hand movements used to provide emphasis, 2) emblems, carrying conventional meanings, 3) deictic gestures, pointing at referents, and 4) iconic and metaphoric gestures, representing concrete and abstract notions.

The qualitative method was then integrated with quantitative research. For the analysis of adjectives, this study employed two corpus linguistics tools, namely:

- Free CLAWS WWW tagger, which is a free web tagging service offered by UCREL at Lancaster University (Garside 1987) and available at http://ucrel.lancs.ac.uk/claws/trial.html.
- AntConc, which is a freeware corpus analysis toolkit for concordances and text analysis developed by Laurence Anthony (Anthony 2016) and available at http://www.antlab.sci.waseda.ac.jp/antconc

In order to proceed to the analysis of adjectives, the corpus was tagged using Free CLAWS WWW tagger, which divides adjectives into 'JJ' (Adjectives), 'JJR' (Comparative Adjectives), and 'JJT' (Superlative Adjectives). The list obtained was then manually cleaned to verify the correctness of the results. For instance, nouns with an adjectival function, such as *key* in *key competencies*, were excluded from the analysis. Based on this analysis, the final total number of occurrences of the tags JJ, JJR, and JJT in the corpus was 1,480. In other words, 4.18% of the tokens in the entire corpus were adjectives. In order to classify these adjectives, Kerbrat-Orecchioni's (1980) and Felices Lago's (1997) taxonomies were used, as explained in section 5.

4. Qualitative analysis: Three case studies

In Genre Theory, Bhatia (2012, p. 24) distinguishes between "intertextuality", which operates within the "textual space" or across "text-internal resources", and "interdiscursivity", which operates across "text-external semiotic resources". The following analysis focuses on both text-internal verbal cues and text-external non-verbal cues.

4.1. An economics Google Talk

The first Google Talk selected for the analysis belongs to the economics sphere and is related to the topic of "Superbosses". The extract in (1) illustrates some of the interactive features and forms of audience engagement used by Sydney Finkelstein in his talk:

(1) It's good to have a chance to share some ideas around superbosses. And I'm really interested in what you have to say about that. And I'm going to have some questions for you as well. A professor can't help that type of stuff.
So let's start with – I'm a foodie. I'm really into great restaurants. That's why I live in Hanover, New Hampshire, for example.
And one of the things I noticed in the early days of getting interested a little bit about talent is I heard about this woman that, after graduating from college, she went to France. And she lived the Paris – the Parisian lifestyle. She went to the markets. […]
And she comes back to America after a few years. […] And she opened up a restaurant in Oakland, in Berkeley, by the name of Chez Panisse.
And some of you will know who this is. It's Alice Waters, legendary chef, who, along with being this innovator when it comes to food, also has, over the course of her career – and I'll say so far because she's still going strong – has spawned literally over 200 people. […]
And I said, wow, here's somebody who's really kind of interesting, kind of cool – and what she has been able to do. (S. Finkelstein, 00:12-01:59)

In this excerpt, the speaker uses the first person pronoun 'I' to introduce himself, his specialised topic, and his personal experience (*<u>I</u>'m really interested*, *<u>I</u>'m going to have some questions*, *<u>I</u>'m a foodie. <u>I</u>'m really into great restaurants*) and the second person pronoun 'you' to address directly to the audience (*what <u>you</u> have to say*, *some questions for <u>you</u>*, *And some of <u>you</u> will know*). The use of these pronouns lowers the level of discourse to familiarity, thus helping increase proximity with the public and create a sense of inclusivity which is comparable to the interactional approach identified in academic lectures by Fortanet (2004) and Walsh (2004) by means of the same devices.

Other verbal devices that are used by the speaker to increase proximity with his audience consist of instances of informal register: i.e., contractions (*It's*, *I'm*, *let's*, *That's*), informal or slang words (*stuff*, *foodie*, *cool*), general words (*one of the things*, *somebody*), and idioms (*she's still going strong*).

Humour is also used as a sort of icebreaker (*A professor can't help that type of stuff*, *That's why I live in Hanover*). It catches the hearers' attention and simultaneously reduces the distance between the expert speaker and the non-expert (or not necessarily expert) audience. Metaphorical language (*she […] has <u>spawned</u> literally over 200 people*) also contributes to the creation of a familiar environment. The verb *spawn* is here used in the sense of 'to lead somebody to success' and referred to employees starting their successful career.

As for the lexicon, specialised vocabulary is not highly technical in the talk, but easily accessible to the layman (e.g. *career*). Furthermore, non-verbal devices are employed to reinforce meaning and strengthen the message conveyed. For instance, the speaker's hands going up and body movement from the left to the right side when saying *over the course of her career* are representative of Alice Waters' career path and success achievement (see Figure 2; cf. 'iconic and metaphoric gestures' in McNeill 1992; for the interpretation of gestures in TED Talks, see Masi 2016).

Figure 2
Example of hand gesture and body movement in the economics Google Talk.

Finally, what distinguishes Google Talks from the textual viewpoint is the use of narration, prevailing over argumentation and instruction. For instance, the speaker in (1) exemplifies an abstract concept (i.e. *superbosses*) by narrating the actual story of Alice Waters (*I heard about this woman that, after graduating from college, she went to France, And she opened up a restaurant in Oakland*). Concrete experiences and actual actions can assist the audience in understanding abstract concepts or ideas, such as 'superbosses' and 'successful career'.

4.2. A political Google Talk

The second case study is a political Google Talk on "Naked Diplomacy", based on the homonymous book by Thomas Fletcher.

Extract (2) especially shows the interplay of different forms of humour, narration, and gestures:

(2) There's a story about when my son turned up at the PM's house as a four-year-old completely naked, and the housekeeper came to the door and said, you're the first guest the Prime Minister's ever had arrived naked. A sign of how professional the British Foreign Service is.
There's a story in there about the time I was at the G8 Summit with David Cameron – his very first summit. And I don't know if it's like this in your world, but there's a kind of physicality and theatricality to the way that leaders interact. […]
Word went round, and when we arrived at the meeting, Obama was feeling his muscles, and Sarkozy was looking a bit jealous, and so on. Perfect.

> And then Berlusconi saw this and was a bit put out that someone else was stealing that space. And so he disappeared off and came back 10 minutes later with a stack of photos of himself lightly oiled and wearing a pair of very, very tight Speedos. (T. Fletcher, 03:59-05:10)

Humour occurs here in the form of incongruity: there is a contrast between the absurdity of the situation (*my son turned up at the PM's house as a four-year-old completely naked*) and the cold reaction of the housekeeper (*the housekeeper came to the door and said, you're the first guest the Prime Minister's ever had arrived naked*).

Still another form of humour is addressed towards politicians attending the G8 Summit. In particular, three politicians are targeted by the speaker, namely, the US ex-President Barack Obama (*Obama was feeling his muscles*), the French ex-President Nicolas Sarkozy (*Sarkozy was looking a bit jealous*), and the Italian ex-Prime Minister Silvio Berlusconi (*then Berlusconi saw this and was a bit put out*). Figures 3 shows the speaker's use of body posture and arm gestures jokingly imitating Obama's attitude.

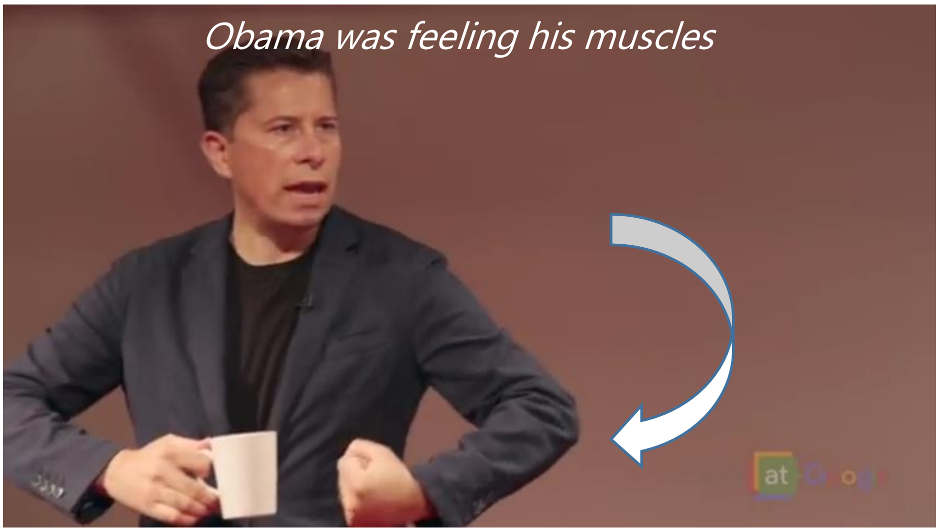


Figure 3
Example of body posture and arm gestures in the political Google Talk.

By contrast, Berlusconi was especially ridiculed by the photographs that he showed of himself wearing a Speedo swimsuit (*photos of himself lightly oiled and wearing a pair of very, very tight Speedos*) and by the metaphor referring to the other politicians attracting the attention in his place (*someone else was stealing that space*) (cf. Scotto di Carlo 2013 and Mattiello 2017 for different forms of humour in TED Talks).

Finally, the prevailing text type in (2) is again narration, as the speaker uses first person pronouns (*I*, *we*) and verbs in the simple past tense to describe his personal experiences (*turned up*, *came*, *said*, *went*, *arrived*, etc.), as a way to share them with the audience. The use of colloquial expressions (*Word went round* 'the news was disseminated', *put out* 'aggrieved', *a stack* *of* ‘a large quantity of’) lastly helps create a familiar atmosphere and increase proximity with the audience.

4.3. A scientific Google Talk

The last Google Talk selected for the qualitative analysis belongs to the scientific field. In this field, many research talks are nowadays presented at conferences using visuals such as slides, transparencies, or PowerPoint displays (Carter-Thomas, Rowley-Jolivet 2003; Hertz 2015), thus allowing speakers to be better understood and arousing higher interest among participants. However, unlike conference presentations, whose language is more prepared and close to written language (Wulff *et al.* 2009), Google Talks are not completely planned speech events, nor have they to comply with the time slot provided or with the specialised tone of a conference.

For instance, the presenter of the Google Talk entitled “Anatomy of a Breakthrough in Targeted Cancer Treatments”, Dr. Brian Druker, developer of a chemotherapy medication used to treat cancer, sounds spontaneous, and his tone is relaxed, even if the topic involves a serious matter. In extract (3), for instance, he is using a rather familiar tone when describing the disease on which he is doing research (i.e. chronic myelogenous leukemia):

(3) Now the disease I worked on – called chronic myeloid leukemia – started with its first description in 1845. We understood the cause by about 1985. And by 2001, we had a specific therapy. So that’s the timeline for the disease I worked on.
The other thing that’s important about this is, this is what I call my translational research slide. [POINTS AT FIRST SLIDE]
You describe a clinical entity. You understand it. And then you can actually do something about it. [SHOWS SECOND SLIDE]
So clinical description of CML. [SHOWS THIRD SLIDE] […]
And just to give you an idea of what the current generation is growing up with – my daughter is eight years old, my youngest daughter. And she just recently achieved a black belt in taekwondo. And as a reward, she wanted me to buy her a typewriter so she could type out old fashioned Google Docs. That was her impression of the world. She has not lived in a world without Google. (B. Druker, 03:42-05:29)

Familiarity is especially created by verbal strategies. Short sentences are introduced by discourse markers (*Now*, *And*, *So*). General words prevail over precise or technical ones (*The other <u>thing</u> that’s important*, *you can actually do <u>something</u> about it*). Specialised abbreviations are only given as anaphoric references to their full forms (e.g. *CML* referring back to *chronic myeloid leukemia*), whereas more colloquial clippings are left unexpanded (e.g. *Google Docs* for ‘documents’), since they are familiar also to non-experts.

Furthermore, the speaker addresses directly his audience by using a second person pronoun ‘you’ (*And just to give <u>you</u> an idea*), thus enhancing proximity with hearers, and repeatedly points at his slideshow to gain the audience’s attention and improve understanding (cf. ‘deictic-spatial gestures’ in McNeill 1992). The speaker's gaze is always directed towards the hearers as an effort to keep eye contact. Intimacy and empathy are also created via gesturing and facial expressions, for instance when Dr. Druker intertwines his hands and adopts a serious expression of concern (see Figure 4).

Figure 4
Example of gesturing and facial expression in the scientific Google Talk.

Descriptive and informative text types, which are typically required in scientific discourse, here co-occur with self-narration (e.g. *my daughter is eight years old, my youngest daughter. And she just recently achieved a black belt in taekwondo. And as a reward, she wanted me to buy her a typewriter*). Finally, the speaker adopts an ironic tone when he comments on his daughter's thought about the 'old-fashioned' search engine Google (*That was her impression of the world. She has not lived in a world without Google*). Besides familiarity, both narration of personal events and humour help release the tensions deriving from the serious topic dealt with.

However, the three features highlighted in (3) are not the only strategies adopted by the speaker to involve hearers and reinforce meanings. Unlike conference presentations (Bucher, Niemann 2012; Hertz 2015; Jurado 2017), Google Talks have no time limits, and this allows speakers to enrich their PowerPoint-supported presentations with additional electronic devices, such as embedded videos. For instance, Dr. Druker includes in his talk the video clip of a young cancer survivor, Katie Knudson, as shown in extract (4):

(4) B.D.: But the reality is, there are lots of people who are benefiting. And let me just share one story with you. [VIDEO PLAYBACK]
K.K.: Hello, everyone. My name is Katie. And I'm very excited to be here tonight. I'm going to start out by telling you all a little bit about myself. I'm 18 years old. I'm a nursing student at University of Portland. I graduated in the top 8% of my high school class. I was… [APPLAUSE] Thank you.
I was part of the Royal Crowns Dance Team, where I was team captain, all state athlete, and state champion. I love my dogs, Italian food, and "Grey's Anatomy".

> And when I was six years old, I was diagnosed with chronic myeloid leukemia.
> We all know what happens in a treatment for cancer – chemo and radiation, hair loss and nausea.
> But would you like to know a secret? I never had any of that. I had Gleevec.
> Because of Brian Druker, my life is as I described it to you. Dr. Druker began developing Gleevec the same year that I was born. And it was FDA approved just one month before I was diagnosed. And if you want to talk about timing, that's some of the best that I've ever had. (B. Druker/Katie Knudson, 21:48-23:06)

The video 'narrates' Katie's story from the teenager's viewpoint. She is not a specialist, of course, but a young woman who has experienced the illness and the treatment and decided to share her experience with Imatinib (or *Gleevec*) to provide evidence of its extraordinary effects. Hence, the language involved is far from being specialised or technical. On the contrary, she uses naïve expressions (*I'm very excited to be here tonight*), simple sentences (*I love my dogs, Italian food, and "Grey's Anatomy"*), and questions (*But would you like to know a secret?*) to involve the audience.

This embedded video demonstrates that, in Google Talks, scientific communication is not only addressed to an audience of specialists, as it generally happens in research talks (Swales 2004), but reframed and recontextualised in a new digital genre accessible also to non-specialists, or even absent addressees: i.e., people who are fighting daily against cancer.

5. Quantitative analysis: The case of evaluative adjectives

Let us turn now to the quantitative corpus-based analysis, which focuses on evaluative adjectives. Evaluative adjectives as a means of expressing epistemic stance have attracted the attention of several scholars in the field of specialised and popularisation discourse. For instance, Caliendo and Compagnone (2014, pp. 119-120) have remarked the co-occurrence of the stance marker *think* with the evaluative adjectives *interesting*, *major*, *best*, and *complicated* in TED Talks, as a way to express a subjective opinion.

In a study on the same genre, Scotto di Carlo (2015) has shown that TEDsters use emotional and evaluative adjectives to transform their presentations into a more personal experience with the audience. In particular, she concludes her article by claiming that, in TED Talks, both aesthetic and emotive adjectives allow "the speakers to convey their knowledge humanising the intellectual experience, getting close to what the audience feels" and "guide the audience to accept [the speakers'] claims" (Scotto di Carlo 2015, p. 214). In general, she concludes that evaluative adjectives are crucial in knowledge dissemination, as they appeal to the audience's sense of identity, self-interest, and emotions. Thus, it is my interest in this study to verify whether or not these adjectives are equally relevant to Google speakers and audience.

My analysis of adjectives in the three Google Talks selected draws upon Kerbrat-Orecchioni's (1980) and Felices Lago's (1997) classifications.

In line with Kerbrat-Orecchioni (1980), adjectives can be distinguished into objective, which enunciate a quality independent from the enunciator (e.g. *single*, *male*), and subjective, which imply an emotive reaction or value judgement (e.g. *happy*, *pathetic*). The subjective class of evaluative adjectives is further divided into axiological and non-axiological. Non-axiological adjectives, such as *hot* or *large*, imply a qualitative or quantitative evaluation of the modified noun, but do not reflect any emotion on the part of the speaker/writer. By contrast, axiological evaluative adjectives, such as *correct* or *nice*, are fully subjective, as they imply a qualitative evaluation, adding a positive or negative judgement to the modified noun. In other words, axiological adjectives reflect the speaker's/writer's favourable or unfavourable position with regard to the modified noun.

In section 5.1, the objective adjectives retrieved in my small corpus were analysed in the three talks separately, in order to establish whether they belonged to general or specialised vocabulary, whereas in section 5.2, given their higher number, the subjective adjectives were analysed in the three talks taken together.

In section 5.2, the quantitative results deriving from an analysis of the subjective adjectives were then sub-categorised into semantic groups. The non-axiological adjectives were classified into the gradable categories of 'quantity' (*multiple*), 'colour' (*black*), 'position' (*distant*), 'dimension' (*little*), 'material'/'consistency' (*soft*), 'time'/'age' (*new*), 'weather'/'temperature' (*cold*), and 'relational' (*normal*). Similarly, axiological adjectives were classified according to Felices Lago's (1997, p. 105) functional scale, dividing them into ten semantic groups: i.e., 'aesthetics' (*beautiful*), 'emotion'/'behaviour' (*exciting*), 'function'/'pragmatism' (*simple*), 'prominence' (*important*), 'intellect' (*interesting*), 'veracity' (*possible*), 'general qualities' (*good*), 'vitality' (*healthy*), 'religion'/'politics'/'ethics' (*moral*), and 'economy'/'material' (*expensive*).

Overall, the analysis of the subjective adjectives in the corpus has not been easy, as it has required additional examination for disambiguation. For instance, collocational patterns were checked in order to discriminate between the non-axiological and the axiological type: e.g., *advanced* in the collocation *advanced disease* was assigned to 'time' (non-axiological), but, in *advanced technologies*, it was assigned to 'prominence' (axiological). The adjective *high* was similarly ambiguous, in that it generally indicates 'position', but in collocational patterns such as *high quality* or *high risk*, it was better classified as indicating 'prominence'. In other words, *high* was considered non-axiological on some occasions, but axiological on others. Moreover, both *naked* (as in *naked diplomacy*) and *transparent* (as in *honest and transparent*) were classified as belonging to ‘politics’, while *warm*, rather than under the ‘temperature’ label, was put under ‘emotion’ when collocating with *relationships*.

5.1. Objective adjectives

Table 1 shows a list of the top ten most frequent objective adjectives in the corpus, ordered by number of occurrences (O), respectively in the economics Google Talk (EC_GT), political science Google Talk (PS_GT), and medical Google Talk (MD_GT). By definition, objective adjectives enunciate a quality independent from the enunciator. For instance, origin adjectives, such as *American*, *British*, or *African*, do not imply any evaluation or emotion from the speaker’s part.

A closer observation of the data has also shown that the objective adjectives in the corpus mostly correspond to specialised terms. This is chiefly evident in the scientific Google Talk, in which objective adjectives such as *immune*, *clinical*, *toxic*, *infectious*, or *genetic* are the most frequently used, but also in the economics talk, with adjectives such as *professional*, *consulting*, and *hired*, as well as in the political one, with *foreign*, *professional*, *political*, *presidential*, and *civil* as recorded examples.

Token EC_GT	O	Token PS_GT	O	Token MD_GT	O
Organic	4	Foreign	11	Immune	23
American	3	Professional	7	Clinical	13
Professional	2	British	6	Toxic	6
Consulting	2	Iranian	3	Infectious	6
Burning	2	Political	2	Genetic	6
Open	2	Islamic	2	Molecular	5
Academic	2	Online	2	Scientific	4
African	2	Natural	2	Environmental	4
Japanese	2	Presidential	1	Human	4
Hired	1	Civil	1	Open	3

Table 1
Ten most frequent objective adjectives in the corpus.

However, objective adjectives represent a low percentage of the overall number of adjectives found in the corpus, namely 21.03% in MD_GT, 14.03% in PS_GT, and 10.19% in EC_GT (or 15.88% in the whole corpus). The remaining subjective adjectives represent the majority of the adjectives in the corpus (1,245 or 84.12%) and they are supposed to contribute to the level of subjectivity of the talks.

5.2. Subjective adjectives

All the subjective adjectives resulting from the analysis of the corpus were manually categorised according to Kerbrat-Orecchioni's (1980) axiological vs. non-axiological distinction. The analysis revealed that a majority of the 1,245 subjective adjectives belonged to the axiological group (i.e. 767, corresponding to 61.60%), while 478 (38.39%) were classifiable as non-axiological.

5.2.1. Non-axiological adjectives

The non-axiological adjectives were categorised into the above-mentioned gradable categories. Table 2 illustrates the ten adjectives with the highest number of occurrences (O) for each category in the entire corpus. For reasons of space, some of the adjectives with the same number of occurrences had to be excluded. The selection was mainly made on the basis of synonymous adjectives already represented in the table: e.g., *intermediate*, *countless* (1 occ.) were included, while their respective synonyms *halfway*, *endless* (1 occ.) were not.

Quantity	O	Colour	O	Position	O	Dimension	O
Single	6	White	2	Higher	4	Big	22
Multiple	4	Black	2	High	2	Biggest	12
Only	3	Blue	1	Far	1	Long	10
Massive	2	Red	1	Near	1	Large	10
Quadruple	2	Gray	1	Spatial	1	Little	7
Much	1			Low	1	Gigantic	6
Scarce	1			Lower	1	Huge	6
Double	1			Intermediate	1	Bigger	5
Countless	1			Bilateral	1	Small	5
Continuous	1			Bottom	1	Three-dimensional	3
Material/ Consistency		**Time/Age**		**Weather/ Temperature**		**Relational**	
Soft	3	New	35	Warm	1	Different	31
Rough	1	Early	24			Whole	16
Granulocytic	1	Old	11			Specific	13
Solid	1	Chronic	10			Social	11
		Traditional	6			Similar	10
		Earlier	6			Entire	8
		Unusual	5			Individual	5
		Younger	5			Related	4
		Youngest	5			Common	4
		Advanced	5			Global	4

Table 2
Ten most frequent non-axiological adjectives in the corpus.

It is not surprising that some of the non-axiological categories, such as 'colour', 'material', and 'weather', are less relevant than others (e.g. 'dimension', 'time', 'relational') in terms of both frequency and range of items that they include. Within 'dimension', it is worth noting that adjectives indicating big size (e.g. *big*/*bigger*/*biggest*, *long*, *large*, *gigantic*, *huge*) prevail over those referring to small size (e.g. *little* or *small*). In the same way, within 'time', adjectives denoting novelty (*new*, *early*/*earlier*, *unusual*, *younger*/*youngest*) prevail over those indicating tradition (*old*, *traditional*). In general, the latter result suggests that the speakers' attitude is growth- and future-orientated.

5.2.2. Axiological adjectives

The axiological adjectives were similarly divided into multilevel categories, or prototypical evaluative terms (Felices Lago 1997, p. 105), as can be seen in Table 3. For reasons of space, only the ten most used adjectives for each category have been reported in the table, using for the selection the same criteria as those adopted for Table 2.

Aesthetics	O	Emotion/ Behaviour	O	Function/ Pragmatism	O	Prominence	O	Intellect	O
Great	26	Crazy	7	Effective	14	Important	18	Interesting	22
Amazing	9	Tremendous	5	Ready	9	Famous	11	Interested	10
Special	6	Competitive	5	Hard	8	Successful	7	Clear	4
Greater	6	Favorite	4	Powerful	7	Head	7	Smart	4
Super	5	Friendly	4	Simple	5	Senior	6	Curious	4
Fascinating	5	OK	3	Difficult	4	Remarkable	4	Brilliant	2
Unbelievable	5	Willing	3	Complicated	4	Louder	4	Analytical	2
Legendary	4	Exciting	3	Tough	4	Prevailing	3	Creative	2
Incredible	4	Loved	3	Adaptable	3	Relevant	3	Obvious	2
Cool	3	Sorry	3	Easy	3	Leading	3	Logical	1
Veracity	O	General Qualities	O	Vitality	O	Religion/Politics/ Ethics	O	Economy	O
Real	18	Good	41	Lethal	7	Right	43	Poor	3
Sure	13	Better	19	Developing	5	Naked	11	Untapped	3
Certain	8	Best	16	Agile	3	Wrong	7	Rich	2
True	7	Bad	8	Fastest	3	Holy	4	Sustainable	2
Possible	3	Optimistic	6	Safer	2	Fair	4	Richer	1
Likely	3	Positive	5	Fatal	2	Diplomatic	3	Inexpensive	1
Authentic	3	Pessimistic	2	Slow	2	Transparent	2	Balanced	1
Genuine	2			Rapid	2	Uncompromising	2	Unbalanced	1
Potential	2			Fast	1	Honest	2	Scalable	1
Supposed	2			Dead	1	Secret	2	Attained	1

Table 3
Ten most frequent axiological adjectives in the corpus.

In general, the high token frequency of adjectives referring to positive qualities in all categories, such as *good*, also in its comparative and superlative forms (76) (vs. *bad* 8), *right* (43) (vs. *wrong* 7), *important* (18), or *interesting* (22), demonstrates that the speakers' overall attitude is optimistic, confident, and encouraging. As a result, the audience will be inspired, even reassured, by the speakers' talks.

A closer observation of the data in the axiological table also shows that 'aesthetics' and 'emotion'/'behaviour' are quite numerous classes of adjectives. In other words, many of the axiological adjectives are used to express aesthetic appreciation and emotive reactions. Moreover, looking at the quantitative data, it can be noticed that most of the adjectives attribute positive aesthetic or emotive properties. Aesthetic axiological adjectives, for instance, include *amazing*, *great*, *fantastic*, *cute*, *super*, and *cool* as relevant examples. The same can be observed for the emotional adjectives *OK*, *favorite*, *friendly*, *exciting*, or *fun*, all rather atypical in specialised (con)texts. The emotions triggered by these adjectives let the audience perceive the speakers as closer to them, and make them feel the same positive emotions.

However, the category of 'emotion'/'behaviour' is much more varied than the others and many 1-occurrence adjectives which had to be excluded from the table (e.g., *abysmal*, *disastrous*, *emotional*, *frightening*, *frivolous*, *gratifying*, *happy*, *horrific*, *hostile*, *mad*, *motivated*, *nuts*, *objectionable*, *pleased*, *proud*, *provocative*, *rancorous*, *rewarding*, *ridiculous*, *shocking*, *shy*, *striking*, *surprising*, *unwilling*, *upset*, *unfriendly*, etc.) can actually testify to the range of emotions/attitudes involved in Google Talks.

Therefore, like TEDsters (Scotto di Carlo 2015), also Google speakers appear to have the overall goal to create a shared emotional experience with the audience. Indeed, they use aesthetic and emotion adjectives to guide the audience to understand their perspective and accept their claims. Aesthetic and emotive adjectives are crucial in knowledge dissemination, in that they appeal to the audience's emotions, thus entailing a higher degree of involvement and emotive participation, which facilitates speaker-hearer proximity, intimacy relationships, and sharing of ideas.

6. Summary and final remarks

This study has shown that Google Talks is a hybrid genre lying at the intersection of a series of more or less traditional oral genres, such as university lectures, conference presentations, and TED Talks. In particular, the study has shown that, given their diverse audience (with different degrees of expertise), Google speakers display a blending of discursive practices from different genres and discourses and harness the affordance of new media to achieve their rhetorical purposes, as well as to reduce the distance with their non-expert audience.

The analysis conducted here has shown that, in Google Talks, specialised vocabulary and technical subjects are made more accessible to laymen or non-specialists via a range of linguistic strategies. First, speakers often adopt an informal register, figurative language, and humorous tone, which may help increase proximity with hearers. Second, they use subjective adjectives to express their stance, opinions, and ideas, or to convince their audience. Another frequently occurring strategy is the use of narratives and anecdotes, or the introduction of specialised subjects by means of personal or other people's experiences. Narration reflects a style that appeals to emotions and is especially useful to encourage participation, or to create a familiar ambiance where the audience may feel at ease and involved.

Therefore, this hitherto unexplored genre displays features of various and different other genres, both traditional and more innovative ones. On the one hand, like academics, Google presenters use an informal language (e.g. contractions, slang words, general words, idioms) and figuration (esp. metaphor) to enhance proximity with the audience. Moreover, in both genres, hand gestures and head/body movements are used to reinforce meanings and to facilitate understanding. However, in Google Talks, interaction is obtained via direct address to receivers ('you') and narration prevails over instruction, thus helping exemplify abstract concepts by using concrete experiences.

Google Talks also share some features with conference presentations. For instance, the use of visual support in PowerPoint presentations is shared by the two genres. Moreover, in Google Talks, a monologic first part is followed by a dialogic Q&A part, with a moderator who fills a role similar to that of a session chair in a conference presentation. However, Google speakers' language is more spontaneous, close to relaxed conversation, with discourse markers, specialised abbreviations that are anaphoric referents to their full forms, and colloquial clippings that are typical of spoken discourse as main verbal strategies. As for non-verbal strategies, Google speakers direct their gaze towards their recipients, and even use embedded videos to support their claims. Therefore, personal experiences may be even narrated by a third party, providing a different viewpoint from the expert's one.

However, TED Talks is the closest genre to Google Talks. These two genres share a reduced technicality in specialised content, the use of narratives, and a humorous tone. Humour occurs in various forms, including a sense of contrast and ridicule addressed to others or irony used to release the tensions deriving from a serious topic. Furthermore, narration and hand gestures are used by both TEDsters and Google speakers to arouse a sense of sympathy or empathy.

The closeness between TED and Google Talks has also been confirmed by a quantitative analysis of evaluative adjectives. Although the quantitative analysis is small-scale and limited to three case studies, the most frequent and relevant evaluative adjectives in the corpus have highlighted their importance in hearer's engagement. In my corpus, subjective adjectives (84.12%) prevail over objective ones, thus contributing to the level of subjectivity of the talks. Moreover, axiological evaluative adjectives mainly belong to the aesthetics and emotion classes, that is, they express aesthetic appreciation and appeal to the audience's emotions. Specifically, aesthetic axiological adjectives (e.g. *amazing*, *great*, *fantastic*) refer to positive properties, while emotional ones (e.g. *OK*, *favorite*, *exciting*) trigger emotions which help KD in a sympathetic or empathetic ambiance.

Therefore, Google Talks can be defined as:

- A hybridised genre taking full advantage of the impact of the Internet and digital technologies on the recipient. The communicative immediacy of the medium and the wide spectrum of multimodal practices offer Google speakers the opportunity to reinforce meaning and facilitate understanding by using non-verbal strategies (e.g. hand/arm gestures, body movements, gaze, facial expressions) and visual support (e.g. images, videos) in their PowerPoint presentations. As a result, in Google Talks, specialised knowledge is recontextualised in a more interactive and dynamic web-based multimodal setting.
- A worldwide popularising form of KD. Indeed, while university lectures are addressed to a disciplinary community of semi-/non-experts (i.e. students) and conference presentations primarily target a community of specialists (mainly academics), Google Talks engage with a wider audience, including expert and professional communities, but also non-specialists or experts in other fields. Therefore, their use of an informal and familiar register, humour, personal narratives, and non-verbal semiotic resources is primarily meant to reduce the asymmetry between expert speaker and non-expert audience.

Thus, from the viewpoint of the genre aims, the goal of Google speakers is not only the transmission of specialised knowledge or the sharing of inspiring ground-breaking ideas, but also the sharing of feelings, thoughts, and (generally positive) personal experiences meant to trigger the hearer's emotive reactions.

Lastly, from the viewpoint of the genre use within the academia, as with other oral genres, the multimodal asset of Google Talks could be exploited in educational contexts (cf. Masi 2016 for the important role of gestures in TED Talks to facilitate memorisation/understanding in foreign language teaching). Given their adaptation to the needs of a wider audience and orientation to interaction and extensive participation on digital platforms, Google Talks particularly suit the requirements of ESP in current university contexts, and could be exploited in educational settings, for example, to improve foreign learners' speaking and listening comprehension skills.

The author's future research intends, first, to verify Google Talks' features on a larger scale, by using a more extensive corpus also involving other domains, and, second, to investigate the linguistic features that more noticeably distinguish Google Talks from TED Talks. A larger corpus of Google Talks could allow us to identify, besides length, a final discussion session, and the usual presence of videos embedded in the talks, the specific linguistic differences between these two genres sharing the same medium, similar goals and audiences.

Bionote: Elisa Mattiello is Researcher in English Linguistics in the Department of Philology, Literature and Linguistics of the University of Pisa. She holds a Ph.D. in English Linguistics from the same University, where she teaches undergraduate courses of ESP. Her main scientific interests are in the fields of English morphology, varieties, and specialised language. In these fields, she has authored four monographs – *An Introduction to English Slang* (2008, Polimetrica), *Extra-grammatical Morphology in English* (2013, De Gruyter), *The Popularisation of Business and Economic English in Online Newspapers* (2015, CSP), *Analogy in Word-formation* (2017, De Gruyter) – and numerous articles. Recently, her research has expanded in the direction of popularisation discourse, with the publication of various articles on digital genres of knowledge dissemination in national and international journals or collections. She is currently part of the Editorial Boards of the *International Journal of English Linguistics* and *Cambridge Scholars Publishing*.

Author's address: elisa.mattiello@unipi.it

Acknowledgements: I wish to thank the three speakers of the Google Talks selected for the analysis – i.e. Sydney Finkelstein (Professor of Management at Dartmouth College), Thomas S.F. Fletcher (former UK Ambassador to Lebanon), and Brian J. Druker (Director of the Knight Cancer Institute, Oregon) – for giving me their kind permission to include some extracts and screenshots from their talks in my paper.

References

Anthony L. 2016, *Looking from the past to the future in ESP through a corpus-based analysis of English for Specific Purposes journal titles*, in "English Teaching and Learning" 40 [4], pp. 91-107.

Artiga León M.R. 2006, *The semantic-pragmatic interface of authorial presence in academic lecturing phraseology*, in "Ibérica" 12, pp. 127-144.

Baldry A. and Thibault P.J. 2006, *Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook*, Equinox, London/Oakville.

Bamford J. 2009, *Patterns of description in lectures in science and technology*, in Radighieri S. and Tucker P. (eds.), *Point of View: Description and Evaluation across Discourses*, Officina Edizioni, Rome, pp. 195-210.

De Beaugrande R. and Dressler W.U. 1981, *Introduction to Text Linguistics*, Longman. London.

Bhatia V.K. 1993, *Analysing Genre: Language Use in Professional Settings*, Longman, London.

Bhatia V.K. 1995, *Genre-mixing and in professional communication: The case of 'private intentions' v. 'socially recognised purposes'*, in Bruthiaux P., Boswood T. and Bertha B. (eds.), *Explorations in English for Professional Communication*, City University of Hong Kong, Hong Kong, pp. 1-19.

Bhatia V.K. 2004, *Worlds of Written Discourse: A Genre-Based View*, Continuum, London.

Bhatia V.K. 2012, *Critical reflections on genre analysis*, in "Ibérica" 24, pp. 17-28.

Bucher H.-J. and Niemann P. 2012, *Visualizing science: The reception of powerpoint presentations*, in "Visual Communication" 11 [3], pp. 283-306.

Caliendo G. 2012, *The popularization of science in web-based genres*, in Caliendo G. and Bongo G. (eds.), *The Language of Popularization: Theoretical and Descriptive Models / Die Sprache der Popularisierung: theoretische und deskriptive Modelle*, Peter Lang, Bern, pp. 101-132.

Caliendo G. and Compagnone A. 2014, *Expressing epistemic stance in university lectures and TED talks: A contrastive corpus-based analysis*, in "Lingue e Linguaggi" 11, pp. 105-122.

Calsamiglia H. and van Dijk T.A. 2004, *Popularization discourse and knowledge about the genome*, in "Discourse & Society" 15 [4], pp. 369-389.

Carter-Thomas S. and Rowley-Jolivet E. 2003, *Analysing the scientific conference presentation (CP). A methodological overview of a multimodal genre*, in "ASp" 39-40, pp. 59-72.

Compagnone A. 2014, *Knowledge dissemination and environmentalism: Exploring the language of TED Talks*, in Chiavetta E., Sciarrino S. and Williams C. (eds.), *Popularisation and the Media*, Edipuglia, Bari, pp. 7-25.

Crawford Camiciottoli B.B. 2008, *Interaction in academic lectures vs. written text materials: The case of questions*, in "Journal of Pragmatics" 40, pp. 1216-1231.

Crawford Camiciottoli B.B. 2015, *Elaborating explanations during Open CourseWare humanities lectures: The interplay of verbal and nonverbal strategies*, in Crawford Camiciottoli B.B. and Fortanet-Gómez I. (eds.), *Multimodal Analysis in Academic Settings. From Research to Teaching*, Routledge, New York, pp. 144-170.

Crawford Camiciottoli B.B. 2016, *Chapter Three: A multimodal analysis of interaction in academic lectures: A case study*, in Bonsignori V. and Crawford Camiciottoli B.B.

(eds.), *Multimodality across Communicative Settings, Discourse Domains and Genres*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 65-92.
Fairclough N. 2003, *Analysing Discourse: Textual Analysis for Social Research*, Routledge, London/New York.
Felices Lago Á. 1997, *The integration of the axiological classeme in an adjectival lexicon based on functional-lexematic principles*, in Butler C.S., Connolly J.H., Gatward R.A. and Vismans R.M. (eds.), *A Fund of Ideas: Recent Developments in Functional Grammar*, IFOTT, Amsterdam, pp. 95-112.
Fortanet I. 2004, *The use of 'we' in university lectures: reference and function*, in "English for Specific Purposes" 23, pp. 45-66.
Garside R. 1987, *The CLAWS Word-tagging System*, in Garside R., Leech G. and Sampson G. (eds.), *The Computational Analysis of English: A Corpus-based Approach*, Longman, London, pp. 30-41.
Garzone G. 2006, *Perspectives on ESP and Popularization*, CUEM, Milano.
Garzone G. 2012, *Where do Web genres come from? The case of blogs*, in Campagna S., Garzone G., Ilie C. and Rowley-Jolivet E. (eds.), *Evolving Genres in Web-mediated Communication*, Peter Lang, Bern, pp. 217-242.
Gotti M. 2014, *Reformulation and recontextualization in popularization discourse*, in "Ibérica" 27, pp. 15-34.
Hertz B. 2015, *Spotlight on the Presenter. A Study into Presentations of Conference Papers with PowerPoint*, PhD thesis, Wageningen University, Wageningen.
Hood S. and Forey G. 2005, *Introducing a conference paper: Getting interpersonal with your audience*, in "Journal of English for Academic Purposes" 4 [4], pp. 291-306.
Hyland K. 2010, *Constructing proximity: Relating to readers in popular and professional science*, in "Journal of English for Academic Purposes" 9, pp. 116-127.
Jurado J.V. 2017, *A Multimodal Approach to Persuasion in Oral Presentations*, PhD thesis, Universitat Jaume I and Ghent University, Spain.
Kress G. and van Leeuwen T. 1996, *Reading Images. The Grammar of Visual Design*, Routledge, London.
Kerbrat-Orecchioni C. 1980, *L'enonciation de la subjectivité dans le langage*, Armand Colin, Paris.
Luzón M.J. 2013, *Public communication of science in blogs: Recontextualizing scientific discourse for a diversified audience*, in "Written Communication" 30 [4], pp. 428-457.
Masi S. 2016, *Gestures in motion in TED Talks: Towards multimodal literacy*, in Bonsignori V. and Crawford Camiciottoli B.B. (eds.), *Multimodality across Communicative Settings, Discourse Domains and Genres*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 146-165.
Mattiello E. 2017, *The popularisation of science via TED Talks*, in "International Journal of Language Studies" 11 [4], pp. 77-106.
Mauranen A. 2013, *Hybridism, edutainment, and doubt: Science blogging finding its feet*, in "Nordic Journal of English Studies" 13 [1], pp. 7-36.
McNeill D. 1992, *Hand and Mind: What the Hands Reveal about Thought*, University of Chicago Press, Chicago.
Myers G. 2010, *The Discourse of Blogs and Wikis*, Continuum, London.
Rowley-Jolivet E. 1999, *The pivotal role of conference papers in the network of scientific communication*, in "ASp" 23-26, pp. 179-196.
Rowley-Jolivet E. and Carter-Thomas S. 2005, *Genre awareness and rhetorical appropriacy: Manipulation of information structure by NS and NNS scientists in the*

International Conference setting, in “English for Specific Purposes” 24, pp. 41-64.

Scotto di Carlo G. 2013, *Humour in popularisation: Analysis of humour-related laughter in TED talks*, in “European Journal of Humour Research” 1 [4], pp. 81-93. https://europeanjournalofhumour.org/index.php/ejhr/article/view/43/Scotto%20di%20Carlo (15.08.2018).

Scotto di Carlo G. 2015, *Stance in TED talks: Strategic use of subjective adjectives in online popularisation*, in “Ibérica” 29 [1], pp. 201-222.

Swales J.M. 1990, *Genre Analysis, English in Academic and Research Settings*, Cambridge University Press, Cambridge.

Swales J.M. 2004, *Research Genres: Explorations and Application*, Cambridge University Press, Cambridge.

Talks at Google. Where Great Minds Meet. https://talksat.withgoogle.com/ (10.05.2017).

Walsh P. 2004, *A complex interplay of choices: First and second person pronouns in university lectures*, in Bamford J. and Anderson L. (eds.), *Evaluation in Oral and Written Academic Discourse*, Officina Edizioni, Rome, pp. 32-52.

Wulff S., Swales J. and Keller K. 2009, *‘We have about seven minutes for questions’: The discussion sessions from a specialized conference*, in “English for Specific Purposes” 28 [2], pp. 79-92.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 383-406
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p383

HOW GESTURES CONTRIBUTE TO THE MEANINGS OF IDIOMATIC EXPRESSIONS AND PHRASAL VERBS IN TV BROADCAST INTERVIEWS
A multimodal analysis

GIANMARCO VIGNOZZI
UNIVERSITÀ DI PISA

Abstract – This study aims at shedding light on the representation of some peculiar indicators of spokenness (i.e. idiomatic expressions and phrasal verbs) across TV interviews featuring different interviewees: politicians, business people and personalities from showbiz. More precisely, the purpose of this work is to observe how, to what extent, and with which function these traits, which are more typical of oral and informal language, permeate TV interviews pertaining to different specialized knowledge domains. In order to develop a deeper understanding of the usage and the meaning-making of idiomatic expressions and phrasal verbs in TV interviews, some multimodal aspects will be considered throughout the analysis. Drawing from the findings, it was noticed that these complex indicators of spoken style are abundant in the language of both interviewers and interviewees across all three domains, where they seem to work as lubricants that help smooth interaction and thus engage the audience. Furthermore, gestures with arms and hands appeared to occur together with idiomatic expressions and phrasal verbs quite consistently, often pointing to their figurative meanings.

Keywords: TV interviews; idiomatic expressions; phrasal verbs; multimodality; gestures.

1. Introduction: The interview on television

This paper investigates variation in the usage of two frequent markers of orality in English i.e. idiomatic expressions and phrasal verbs (Biber *et al*. 1999), across TV broadcast interviews concerning different knowledge and professional domains, namely politics, business and economics, and showbiz. The analysis combines quantitative and qualitative methods and also takes advantage of a multimodal approach to account for both the multisemiotic nature of communication through the TV format and the stratified meaning that characterizes idiomatic expressions and phrasal verbs.

In today's society, all kinds of TV-mediated interactions have undoubtedly established themselves as cultural realities for larger and larger audiences. Indeed, their societal impact can be, and often is, on a much larger scale than everyday casual interactions, touching the lives of an ever-growing number of people. Within the broad scenario of broadcast interactions, and screen-mediated communications in general, interviews aired on television are probably one of the most widely used, best-developed and long-lasting formats for disseminating information, which already existed before the invention of television (Ekström 2001). Their great success can be ascribed, according to Clayman and Heritage (2002, p. 29), to the mutual interest between those who are interviewed (normally public figures) and journalists/anchors: "journalists need access to public figures for their livelihood, while public figures need journalists to gain access to what Margaret Thatcher once called 'the oxygen of publicity'". Such a symbiosis leads to the stipulation of an 'implicit contract', where journalists grant public figures access to their TV show in exchange for news-making content that will attract the viewers at home.

What mainly characterizes TV interviews is that the final target of the exchange is the 'overhearing audience' (Heritage 1985), i.e. a non-active participant in the conversation and an ideal outsider from the discourse community represented in the interview. In other words, it is especially for their consumption that the interview is conceived and prepared. As Montgomery (2007, p. 260) states, "interviewers and interviewees know that what they say will be appraised not just by their immediate interlocutor but by who-knows-how-many beyond." This triangular communicative frame (among the interviewer, the interviewee, and the home audience) inevitably shapes the way in which knowledge is constructed during the interview (Furkó, Abuczki 2014). In fact, the *raison d'être* of the TV interview is an underlying asymmetry of knowledge where the speaker who asks questions (i.e. the interviewer) claims a 'lack of knowledge' in the name of a third party (i.e. the viewers at home). Simultaneously, it is taken for granted that the direct recipient (i.e. the interviewee) is informed about the topic of the question. The asymmetry of the situation is also reinforced by the differential rights and responsibilities to ask questions of the participants and, thus, it is closely linked to the "situation-specific institutional identities of the interviewer and the interviewee" (Lauerbach 2006, p. 197). To sum up, the rigid question-driven format of the interaction during the interview and the prescribed roles of the participants lead to the creation of an asymmetrical, rather constrained, and unique dialogic event.

Despite the highly repetitive and codified format, when it comes to the range of topics that they may touch on, TV interviews are a highly versatile genre. Most often they are aired within a TV news programme (news interviews) or within thematic TV shows, which are programmes revolving around a certain knowledge domain, such as a political arena, a business talk show, etc. Indeed, interviews can either be a more or less marginal part of the show (e.g. an interview with a doctor in a medical docu-series) or they can be its most important element (e.g. *Hard Talk*, a BBC programme dedicated to in-depth one-to-one interviews). Regardless of the nature of the TV show, on the basis of the specialized and professional community that the interviewee is representing, we can distinguish between different typologies of interviews such as business interviews, political interviews, celebrity interviews, sports interviews.

From the point of view of language use, it is pivotal to always keep in mind that broadcast interviews are generally products that cater to large audiences and have to meet special requirements regarding timing, structure, accessibility (i.e. through popularization strategies) and, last but not least, linguistic register. Especially in the matter of register, some studies focusing on the language of TV political interviews (Fairclough 1998; Bruti 2016, among others) have defined these interactions as stylistically hybrid, mixing elements typical of different registers (e.g. formal, institutional, informal, colloquial) and discourses (e.g. planned, unplanned, spoken and written). More generally, the process through which language undergoes an adjustment for the benefit of the home audience leans toward the so-called 'conversationalization of discourse.' (cf. Fairclough 1998, 2000). This means that the style becomes increasingly colloquial, while involving emotional and more subjective linguistic strategies that help to build rapport among interlocutors (direct or indirect ones) (cf. Fairclough 1998, 2000). Such a trend actually reflects the growing demand for rapport with the interlocutors by the viewers, especially in our era characterized by the advent of online interactions, where everyone at home has the chance to talk 'directly' to public figures (Sindoni 2013). As a consequence, TV interviews are becoming more engaging and intriguing, representing a form of 'infotainment' (Brants, Neijens 1998, p. 315): they inform the audience about something (such as a specific topic) while they also entertain.

The slant towards colloquialism that tends to characterize TV interviews described above is the focus of this paper. More precisely, the present research investigates how and to what extent some recurrent indicators of spoken, involved, and informal style permeate specialized TV interviews representing different discourse domains. The linguistic phenomena of interest in the study are phrasal verbs and idiomatic expressions, as two separate, but similar, instances of formulaic language that are very frequent in the English language (Biber *et al*. 1999). The specialized domains that frame the interviews analysed here are basically three: political science, business and economics, and showbiz. It is important to highlight that not only do these three discourse domains represent different areas of knowledge and expertise, but they also testify to different linguistic registers. In fact, political discourse would typically call for a formal and politer language whereas in the business and economics world, technicality, straightforwardness, and informality are foregrounded (cf. Crawford Camiciottoli 2007). Showbiz, instead, is the least specialized and constrained knowledge domain among the three, in which informality and casualness are freely admitted. Given the fact that the TV interview is inherently a multimodal and multisemiotic text, in which meaning is created through the intersection of visual elements, verbal language, gestures, and other semiotic cues, the ensuing analysis also takes a multimodal approach which aims to show if and how the stratified meaning of idiomatic expressions and phrasal verbs is reflected in communicative modes other than the verbal. Therefore, the research questions addressed in this paper are: how and to what extent are idiomatic expressions and phrasal verbs used in political, business and showbiz interviews? How is their meaning reflected in co-occurring semiotic stimuli other than the verbal?

2. Idiomatic expressions and phrasal verbs: A brief overview

Idiomatic expressions and phrasal verbs are natural linguistic phenomena in English accompanying native-like linguistic competence. As Searle (1979, p. 50) advocates, speaking idiomatically is so pervasive in the English language that speakers seem to follow an implicit rule: “speak idiomatically unless there is some special reason not to”. Some scholars have also attempted to provide the numbers for this phenomenon. Jackendoff (1995), for example, claims that there are as many fixed expressions as there are words in English. Gardner and Davies (2007, p. 347), referring to phrasal verbs, in particular, maintain that “learners will encounter, on average, one phrasal verb in every 150 words of English”.

The idea of studying idiomatic expressions and phrasal verbs together stems from their intrinsically related and sometimes comparable phrasal nature. As Sinclair (1991) first put it, what sets both linguistic phenomena apart from ‘plain language’ is that they both abide by what he calls ‘the idiom principle’, according to which texts are composed for about 80% of multi-word expressions that constitute single choices in the mental lexicon. Hence, idiomatic expressions and phrasal verbs are some possible outcomes of these single choices, being phrasal semantic units in which meaning cannot be limited to the single lexeme (Sinclair 1996).

More precisely, the point Sinclair (1996, 2008) wants to make is that the meaning and the function of each component of a phrasal unit is not, or not simply, a property of the single word itself, but of the whole phrase. His proposal is to store expressions in which a single semantic and pragmatic choice involves more than one word, like idiomatic expressions, proverbs, clichés and phrasal verbs, in the lexicon together with lexical compounds. Indeed, in all these expressions the constituting elements seem to have lost their semantic identity in favour of a non-compositional meaning.

The concept of 'single choices' (i.e. of holist expressions) would become, over a decade later following valuable advances in the field of psycholinguistics (cf. Cooper 1998, Wood 2010, Wray 2002, among others), one of the pivotal points giving evidence to the existence of a high level of formulaicity in language production, especially, but not only, in spoken interactions.

Within the category of formulaic expressions, which encompasses proverbs, conversational routines, collocations, etc., the main peculiarity of idiomatic expressions and phrasal verbs is that they tend to entail figuration, conventionality and, to some extent, flexibility (Liu 2003, Nunberg *et al*. 1994). As for figuration, it is undeniable that the meaning of idiomatic expressions and of the majority of phrasal verbs is not totally compositional. In other words, a phrasal verb is not the exact sum of the meanings of the single words of which it is composed. The fact that idiomatic expressions and phrasal verbs are highly conventionalized is another critical point that distinguishes them from literal plain language. It is thanks to this feature that their meaning or use is somewhat unpredictable unless the expression is lexicalized as a whole in the speaker's lexicon. Finally, concerning flexibility, we can observe that both of the linguistic phenomena under study tend not to change their canonical forms, although studies have demonstrated that they may undergo some alterations and still maintain their figurative meanings (Nunberg *et al*. 1994).

Building on the discussion above, in this research the label 'idiomatic expressions' is broadly used to indicate some institutionalized and mostly fixed expressions "whose overall meaning does not correspond to the combined meanings of its components" (Philip 2007, p. 1). Thus, they differ from free combinations of words because, in that particular context, they work as fixed non-compositional strings that acquire a figurative meaning. In the same vein, the definition of phrasal verbs used in the present study characterises them as "combinations between a lexical verb and a morphologically invariable particle which functions as a single syntactic unit" (Quirk *et al*. 1985): they are two- or three-word sequences made from a verb that colligates with adverbial or prepositional particles and that also semantically collocates (Baicchi, Rosca 2016). More specifically, this work draws upon Celce-Murcia and Farsen-Freeman (1999), who identify three main kinds of phrasal verbs, i.e. literal, idiomatic and aspectual phrasal verbs. The study carried out here looks at the idiomatic ones,[1] which could also be described as a specific subcategory within idiomatic expressions for their semantically opaque meaning and their rather fixed nature, given the fact that if the particle is removed, their meaning radically changes.

Therefore, the choice of studying phrasal verbs and idiomatic expressions in TV interviews stems from the fact that these dialogic interactions are well

[1] The general expression 'phrasal verbs' will be used throughout the paper, even though it only refers to the idiomatic ones.

known as rather extemporaneous, generally presenting more marked similarities with spoken English. For this reason, they are a privileged site for the occurrence of phraseological and more colloquial expressions (Milizia 2013).

3. Data and methods of analysis

The dataset under analysis is part of the materials used for "The ESP Video Clip Corpus", a multimodal corpus currently being compiled by a group of scholars at the University of Pisa for an inter-university project.[2] The corpus encompasses 216 video clips (and their corresponding transcripts) spanning across various knowledge domains and textual genres. This study revolves around TV interviews, one of the six textual genres represented in the corpus, featuring interviewees with different backgrounds and roles, who were interviewed to discuss topics related to their specific knowledge domain and area of expertise. More precisely, the interviewees involved in this study come from the world of *i*) political science, *ii*) business and economics, and *iii*) showbusiness (showbiz). The following table (1) details the dataset by supplying specific information for each item, e.g. the knowledge domain represented by each interview (first column), the name of the person being interviewed and his/her expertise/occupation (second column), the date on which the interview aired (third column), its length (fourth column), and finally the TV show in which the interview appeared (fifth column).

Domain	Interview title	Date	Length	Source
Political science	Interview with Hillary Clinton (Politician)	13-10-17	20:13 minutes	Channel 4 News
Political science	Interview with Donald Trump (USA President)	25-01-17	20:17 minutes	ABC News Tonight
Political science	Interview with Bernie Sanders (Politician)	13-10-15	24:00 minutes	Meet the Press
Business and economics	Interview with Meg Whitman (HP CEO)	03-11-15	21:33 minutes	Charlie Rose
Business and economics	Interview with Nick Woodman (GoPro CEO)	06-09-14	18:40 minutes	Charlie Rose
Business and economics	Interview with Ursula Burns (XEROX CEO)	01-02-17	23:31 minutes	Charlie Rose
Showbiz	Interview with Tom Hanks (Actor)	22-01-14	22:31 minutes	David Letterman Show
Showbiz	Interview with Sofia Coppola (filmmaker)	05-08-17	20:07 minutes	Charlie Rose
Showbiz	Interview with Sheryl Crow (Singer)	09-09-17	20:01 minutes	Charlie Rose

Table 1
Dataset used for the present study.

[2] This research has been financed by the Italian Ministry for the University (PRIN 2015 no. 2015TJ8ZAS).

All the selected interviews were aired on American television programmes of various kinds, and they all feature American interviewees. In accordance with the principles of data harmonization and balance (Freddi 2013) for building a specialized dataset, the full interviews (video and transcripts) were selected amongst those used for the creation of "The ESP Video Clip Corpus" on the basis of their topic (three interviews per discourse domain), their length (between eighteen and twenty-four minutes of conversation) and, since the analysis is synchronic, the date on which they were broadcast (from 2014 to 2017). Altogether, 64:30 minutes of political interviews, 63:40 minutes of business interviews and 62:30 minutes of showbiz interviews were gathered, orthographically transcribed and then analysed.

The methodological apparatus used for the present research attempts to combine quantitative and qualitative considerations and is especially catered towards the multimodal nature of the corpus. Given the complexity of an analytical approach which takes into account different levels of communication (e.g. verbal and nonverbal), I referred to Bednarek and Caple's (2017) topology for situating research to decide how to conduct this study.

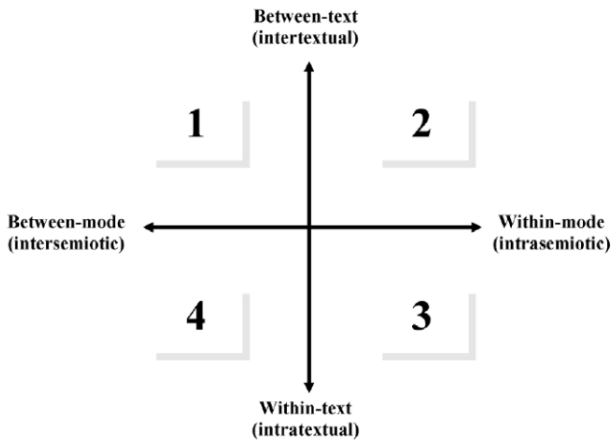


Figure 1
Topology for situating research (Bednarek, Caple 2017).

With reference to the figure above, the analysis was structured into three main phases. The first phase was to detect and transcribe all instances of idiomatic expressions and phrasal verbs in the interviews presented above, together with their time frame of occurrence. The starting point for disambiguation was the definition of idiomatic expressions and phrasal verbs explained in section (2). Phase one could be situated in Zone 1 of the topology, since it entailed the mono-modal (intrasemiotic) analysis of idiomatic expressions and phrasal verbs within each text (and not across texts, as is usually done in corpus-based studies). Preliminary quantitative counts of the frequency of idiomatic expressions and phrasal verbs[3] through a manual check[4] of the transcripts allowed me to explore variations across different discourse domains. The second phase consisted in combining the reading of the transcripts with the observation of the corresponding videos so as to detect whether, how and to what extent the non-compositional meaning of idiomatic expressions and phrasal verbs is also recalled by co-occurring nonverbal semiotic resources (cf. Wildfeuer 2013 for an overview of nonverbal items on screen). Since gestures were found to be the semiotic element whose relationship with the meaning of idiomatic expressions and phrasal verbs was more manifestly evident and recurrent, the second phase of the analysis primarily focused on their description and investigation. As far as their classification and analysis are concerned, I used the taxonomies for gestures put forward by McNeill (1992, 2005) and Kendon (2004). Going back to the topology, in this phase of the research a shift from Zone 1 to Zone 4 can be detected, since the analysis becomes intersemiotic, involving more than just one mode of communication. Thus, still in Zone 4 of the topology, the third phase entailed establishing to what extent and how gestures accompany idioms and phrasal verbs across interviews featuring specialized discourses. For this purpose, a selection of examples taken from different interviews in the dataset under analysis were transcribed multimodally, following a model of transcription already used in Vignozzi (2016), which was adapted from Wildfeuer (2013) and Bruti (2015). The results of this multimodally-informed comparative analysis made it possible to posit some hypotheses about the potential motivation behind the trends detected.

4. Results and discussion

4.1. Quantitative analysis of idiomatic expressions and phrasal verbs

The first step in the analysis consisted in qualitatively going through the transcript of each interview in the corpus in order to single out all instances of idiomatic expressions and phrasal verbs contained in the delivery of the interactants. Table (2) summarizes the results of this preliminary quantitative

[3] As for the identification of idiomatic expressions and phrasal verbs, I relied upon the criteria described in detail in Section 2.

[4] The choice of carrying out a manual check of the transcripts derived from the fact that the linguistic phenomena under analysis in this work are essentially pragmatic in nature. Therefore, an automatic investigation would not have been as accurate, especially in the recognition of idiomatic meanings for which there is a clear form-function mismatch (Aijmer, Rühlemann 2014). Such an attentive analysis was also possible because the multimodal subcorpus used for this study was rather contained in size.

investigation. In the first column on the left, the discourse domains around which each interview revolves are listed, and their titles are displayed in the second column. The third and the fourth column, instead, show the frequency of idiomatic expressions and phrasal verbs in each interview. Finally, the last column on the right illustrates the total count of idiomatic expressions and phrasal verbs occurring in each discourse domain, displaying also their occurrence calculated per minute.

<table>
<tr><th>Domain</th><th>Interview title</th><th>Idiomatic expressions</th><th>Phrasal verbs</th><th colspan="2">Total counts and frequency per minute</th></tr>
<tr><td>Political science</td><td>Interview with Hillary Clinton (Politician)</td><td>14</td><td>22</td><td colspan="2">99 idiomatic expressions + phrasal verbs
(1.53 x minute)</td></tr>
<tr><td>Political science</td><td>Interview with Donald Trump (USA President)</td><td>11</td><td>22</td><td rowspan="2">32 idiomatic expressions</td><td rowspan="2">67 phrasal verbs</td></tr>
<tr><td>Political science</td><td>Interview with Bernie Sanders (Politician)</td><td>7</td><td>23</td></tr>
<tr><td>Business and economics</td><td>Interview with Meg Whitman (HP CEO)</td><td>18</td><td>13</td><td colspan="2">88 idiomatic expressions + phrasal verbs
(0.64 x minute)</td></tr>
<tr><td>Business and economics</td><td>Interview with Nick Woodman (GoPro CEO)</td><td>8</td><td>9</td><td rowspan="2">41 idiomatic expressions</td><td rowspan="2">47 phrasal verbs</td></tr>
<tr><td>Business and economics</td><td>Interview with Ursula Burns (XEROX CEO)</td><td>15</td><td>25</td></tr>
<tr><td>Showbiz</td><td>Interview with Tom Hanks (Actor)</td><td>7</td><td>21</td><td colspan="2">72 idiomatic expressions + phrasal verbs
(0.41 x minute)</td></tr>
<tr><td>Showbiz</td><td>Interview with Sofia Coppola (filmmaker)</td><td>12</td><td>11</td><td rowspan="2">26 idiomatic expressions</td><td rowspan="2">46 phrasal verbs</td></tr>
<tr><td>Showbiz</td><td>Interview with Sheryl Crow (Singer)</td><td>7</td><td>14</td></tr>
<tr><td colspan="2">TOTAL COUNTS</td><td>99</td><td>140</td><td colspan="2">239 (1.25 x minute)</td></tr>
</table>

Table 2
Frequencies of idiomatic expressions and phrasal verbs.

On a surface level, it can be noticed that all interviews in the corpus contain both idiomatic expressions and phrasal verbs in quite a consistent way. Looking at variations among discourse domains, we can appreciate that interviews with politicians feature the highest concentration of these linguistic phenomena with 99 occurrences, corresponding to 1.53 items per minute. Interviews about business and economics contain 88 occurrences of idiomatic expressions and phrasal verbs with a total count of 0.64 instances per minute. Finally, interviews involving showbiz personalities include 72 occurrences, which correspond to 0.41 idiomatic expressions or phrasal verbs per minute. Across all domains, 1.25 idiomatic expressions and phrasal verbs together (239 instances in total) per minute were uttered. This first result is rather low if it is compared with what Erman and Warren (2000) and Liu (2003) found in spontaneous and non-specialized conversation, where Liu (2003) counted around 4 idiomatic expressions per minute.

Examining the data in more detail, phrasal verbs (e.g. ‘figure it out’) were more numerous than idiomatic expressions (e.g. ‘bird in hand’) (140 phrasal verbs and 99 idiomatic expressions). This finding is in line with Biber *et al.*’s (1999) large-scale description of spoken English, according to which phrasal verbs are far more recurrent than idiomatic expressions. Political science interviews contain, by far, the majority of phrasal verbs (67 instances), which, in business and showbiz interviews, are more or less the same in number (47 times in the former and 46 times in the latter). As for idiomatic expressions, the highest concentration is found in business and economics interviews, which contain 41 examples. Political science interviews follow with 32 occurrences and showbiz interviews with 26. Table (3) shows a snapshot of some phrasal verbs and idiomatic expressions found in each domain.

Domain	Sample of Idiomatic expressions	Sample of Phrasal verbs
Political science	Have second thoughts; walk a different line; get their hands dirty; put blood and treasure; perfect storm; back on track; a home run; like a war zone; shine a bright spotlight.	Eked out; swirling around; send [the press secretary] out; getting [sucked] into; paying down; flips out; warned against; flooded with; shut down; run against.
Business and economics	Pull your socks up; been there done that; have a rhinoceros skin; drink your own Kool-Aid; get into the weeds; bird in hand; hit the road; thrown out on your ear; drive the bus into a road or into a ditch; keeping the power plant going.	Lean in to; lever up; take out; come up with; inch along; lean forward; run under; hammer [something] out; manage across; started out.
Showbiz	A rock and a hard place; a ‘B’ movie; throw down the gauntlet; find your way to him; being hot heads; in the dark; at the height of his game; money changes hand; off their game; hit the road.	Filling in; tracked down; cracks [me] up; taking over; cut off; hung out; miss out; broken up; come together; took off.

Table 3
A sample of some idiomatic expressions and phrasal verbs for each domain.

Therefore, as Table (2) shows, the domain in which there are more idiomatic expressions and phrasal verbs is political science, where phrasal verbs are roughly twice as frequent as idiomatic expressions. Given the fact that such linguistic phenomena, as underlined in section (2), are considered typical features of colloquialism and informality, this trend may appear relatively unexpected. Indeed, politics is considered a rather formal and institutional domain, where speakers talk to large audiences and tend to monitor their speech and use an adequately serious tone, akin to a more written-like style (van Dijk 1997). In business interviews, we can observe that idiomatic expressions (e.g. 'change of pace', 'get the ball rolling', etc.) are more numerous than in political interviews and that they are almost as frequent as phrasal verbs. This tendency is in line with the nature of business communication, which is neither particularly formal nor informal, but is a hybrid mixture of technicality and informality (Crawford Camiciottoli 2007). To conclude, showbiz interviews contain the lowest concentration of idiomatic expressions and phrasal verbs, even though they are manifestly less specialized than political and business interviews, and their more relaxed and light-hearted register should, in theory, have favoured the occurrence of idiomatic expressions and phrasal verbs. In fact, these linguistic phenomena are also recognized as rhetorical tools that embellish language, making it more colourful, creative and intriguing (Wood 2010).

4.2. Multimodal analysis

After determining to what extent idiomatic phrases and phrasal verbs occur across the corpus, I added a multimodal perspective to the analysis (the concept of multimodality used here is broadly inspired from O'Halloran 2010 and Wildfeurer 2013). To put it simply, the analysis thus far centred on dialogues only, i.e. one semiotic system within the aural modality, was complemented with the evaluation of the corresponding visual.

The reasons for and the advantages of carrying out a multimodally-informed analysis are manifold. First and foremost, we cannot disregard the fact that TV interviews are essentially thought of and disseminated as audiovisual texts, which are characterized by an intricate semiotic fabric (Baños *et al*. 2013). Differently from a written text, they utilize various 'sign systems' such as language, visual communication, body language, kinesics, etc. by exploiting different modalities (e.g. visual and aural) (O'Halloran 2010). Moreover, given the stratification of idiomatic meaning-making that describes both linguistic phenomena under analysis, a multimodal approach that takes into account different semiotic codes may show, for example, if and how elements from the visual match and co-occur with their complex figurative meaning. Therefore, all idiomatic phrases and phrasal verbs were carefully looked at in their semiotic integrity (i.e. by watching the videos corresponding to the transcripts) in order to observe which nonverbal elements were more manifestly intertwined with their meanings in the interaction.

Such a preliminary inductive exploration of the "modal ensemble that creates meaning" (Kress 2010, p. 59) revealed that body movements (i.e. gestures), were the nonverbal cue that most frequently recalled the non-compositional meaning of both the linguistic phenomena at stake. In particular, it was noticed that arm and hand gestures co-occurred with phrasal verbs and idiomatic phrases in such a consistent and meaningful way so as to inspire me to carry out a more detailed analysis casting light on the role and functions of gestures in sequences involving idiomatic expressions and phrasal verbs.

4.3. Hand and arm gestures

Gesturing is a vast field of study that embraces different communicative events, lying behind the idea that "speech and movement appear together as manifestations of the same process of utterance" (Kendon 1980, p. 208). Giving a complete description of the phenomenon would be far beyond the scope of this research, for which the notion of gestures was narrowed down to the "motion [of arms and/or hands] that embodies a meaning relatable to the accompanying speech" (Kendon 2004, p. 36). As such, the definition of gesture used here does not include movements that do not accompany speech, nor does it include pantomimes (gestures with the obligatory absence of speech) or emblematic gestures (optional presence of speech), such as the 'ok' sign in North America, or signed language in general.

The taxonomy to classify the types of gestures in my dataset was adapted from McNeill (1992), whose classification distinguishes between four types:

- Deictic gestures are pointing motions to identify, in space or time, an entity under discussion.
- Iconic gestures represent a concrete idea. An example given by McNeill is a speaker who, while retelling a scene from a Sylvester and Tweety Bird cartoon in which Tweety Bird stuffs a bowling ball down a drain pipe on top of Sylvester, stuffs one hand with fingers together to form a fist inside a ring created by the other hand.
- Metaphoric gestures, instead, represent an abstract idea. In general, they resemble something concrete in order to represent something abstract (e.g. pantomiming a spherical shape to represent the idea of wholeness).
- Beats are gestures typically executed as rapid hand flicks, which have no semantic content of their own and are thought to play a role in the conception of the discourse organisation. In fact, they generally occur during the introduction of a new character or to highlight important points in discourse.

Moving on to the functions of gestures, I mainly followed Kendon (2004) who, from a more pragmatic point of view, describes how gestures can give precision to the meaning of utterances. He identifies modal gestures, which express degrees of certainty; performative gestures, whose primary function is to perform a speech act (an example could be dismissing something with a wave of the hand); and parsing, which specifically marks different units within an utterance. Weinberg *et al*. (2013) add three more functions to this repertoire: indexical (when used to indicate a referent), representational (when they represent an object or idea) and, finally, social (i.e. gestures that stress the importance of the message or attempt to involve the audience to a major extent).

4.3.1. Idiomatic expressions, phrasal verbs and gestures

For a broader understanding of the role of gestures in the meaning-making, an assessment of whether and to what extent idiomatic expressions and phrasal verbs coexist in time and with tight synchrony with hand or arm gestures was undertaken. First, idiomatic expressions alone were taken into account. The investigation revealed that among the 99 idiomatic expressions identified in the dataset, 62 (62.2%) occur together with gestures.[5] An example is Hillary Clinton, who says ‘have second thoughts’, meaning a change of opinion, while moving her arms and hands from left to right.

The following figure (2) offers a more detailed breakdown divided per specialized discourse domain. The first bars for each domain illustrate the occurrences of idiomatic expressions, the second ones specify the number of times idiomatic expressions co-occur with gestures.

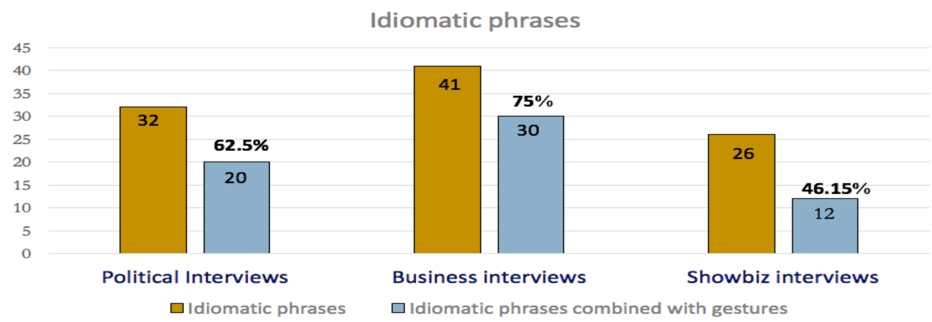


Figure 2
Idiomatic expressions and gestures.

5 In a few cases, it was not possible to evaluate whether the idiom co-occurred with a gesture or not, because the voice was off-camera.

Overall, the results seem to suggest that in business interviews gestures are associated with idiomatic expressions to a major extent (in 75% of the cases). In political interviews as well, gestures are quite preponderant being employed in 62.5% of cases. Finally, in showbiz interviews, they are used in slightly less than the half of the total cases (46.15%).

The same type of analysis was carried out for phrasal verbs, for which in 81 examples out of 160 (50.6%) gestures are intertwined with verbal language. Figure (3) graphically displays the results for each discourse domain.

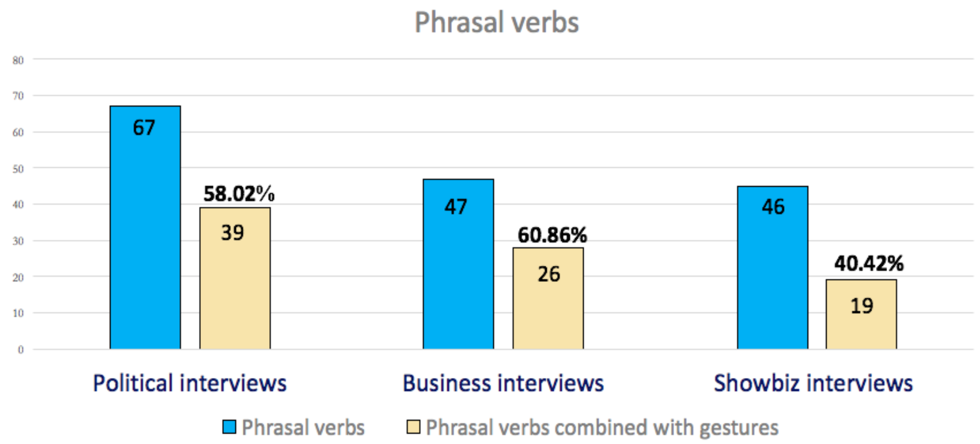
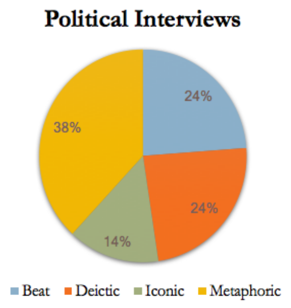


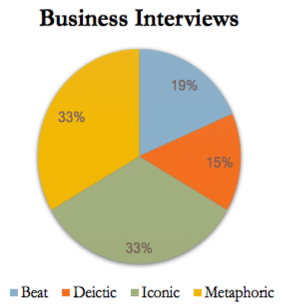
Figure 3
Phrasal verbs and gestures.

Quite interestingly, the results obtained for phrasal verbs show similar trends to those found for idiomatic expressions. In fact, even though these particular verbs are more recurrent in political interviews than in business and economics, the specialized domain with which hand or arm movements are more often associated is again business and economics (60.86%). Moreover, in political interviews gestures appear in 58.02% of cases, and in showbiz interviews, where there is the lowest frequency peak, in just 40.42% of the cases.

4.3.2. Classification of gestures across domains

With the aim of understanding the typology of gestures used in each discourse domain, both idiomatic expressions and phrasal verbs were evaluated following the framework introduced above (cf. section 4.3.). The figures that follow (4, 5, 6) illustrate the analysis for gestures co-occurring with idiomatic expressions. Before describing the results, I would like to point out that, as relevant literature on the topic has noted (cf. Crawford Camiciottoli 2015, Bruti 2016, Masi 2016, among others), hand and arm gestures that recur with verbal language sometimes can be quite hard to classify. In my experience, this is especially the case with metaphorical and iconic gestures, which in some cases may overlap.





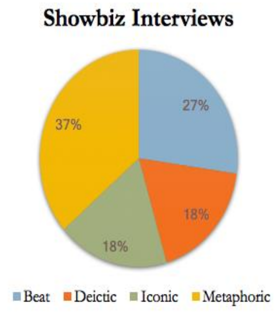
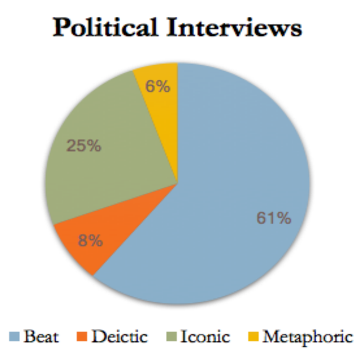


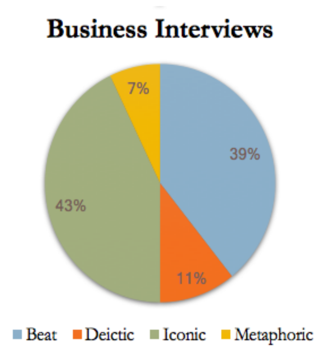
Figure 4 Figure 5 Figure 6

Types of gestures accompanying idiomatic expressions.

What seems to emerge from this analysis is that metaphoric gestures most frequently accompany idiomatic expressions (e.g. the interviewee opening his/her hands and arms while saying 'it's bigger picture') across all the three discourse domains (38% in political interviews, 33% in business interviews and 37% in showbiz interviews). The other type of figurative gestures, i.e. iconic gestures, are also quite recurrent in all interviews, with particular relevance in business interviews (33%), where they are as frequently associated with idiomatic expressions (e.g. the interviewee imitating the act of seizing something while saying 'bird in hand') as occurs with the metaphoric gestures. Looking at beats associated with idiomatic expressions, we can see that their frequency is consistent across the three domains, ranging from 19% in business interviews to 27% in showbiz interviews. To conclude, deictic gestures (e.g. the interviewer saying 'ahead of time' while pointing in front of him), appear to be slightly foregrounded in political interviews (24%), and in business and showbiz interviews they occur to a similar extent (15% and 18%).

The remaining figures below (7, 8, 9) show the types of gestures that are performed in combination with phrasal verbs.





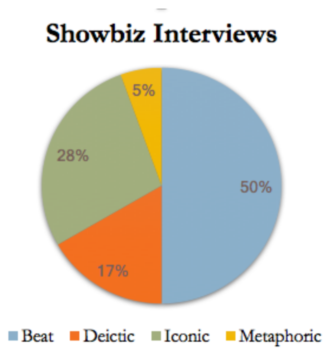


Figure 7 Figure 8 Figure 9

Types of gestures accompanying phrasal verbs.

As can be noticed, beats are the most frequent kind of gestures co-occurring with phrasal verbs. In political interviews in particular, they account for more than half of the total of gestures. As to iconic gestures, they are most frequently employed with phrasal verbs in business interviews (43%), then in showbiz and political interviews to almost the same extent (28% of cases in the former and 25% in the latter). Quite surprisingly, metaphoric gestures, which were the most frequent type with idiomatic expressions, are not very frequently used with phrasal verbs. Finally, deictic gestures are quite stable across the three domains, although, on average, they were employed with a higher frequency with idiomatic expressions.

4.4. Multimodal transcription of selected examples

The subsequent tables illustrate the multimodal transcriptions of a selection of examples showing how co-occurring gestures may contribute to the meaning-making of idiomatic expressions (table 4) and phrasal verbs (table 5). As previously noted, the transcription follows the same framework employed in Vignozzi (2016). The first column on the left gives the title of the interview and the discourse domain. The second column displays the shots that correspond to the uttered expression, which are described in the third column. The fourth is devoted to gestures and is the richest part of the transcription. The last column consists of the transcription of the verbal language making up the idiomatic expression or the phrasal verb.

Interview title and domain	**a) Shot**	**b) Shot description**	**c) Gestures**	**d) Spoken language**
(1) Interview with Senator Bernie Sanders - Political science	 	Medium shots of Bernie Sanders	Hands and arms gesture, closed hand, index fingers pointing, moving from up to down. Arms moving with the hand **Gesture function:** Representati onal/metaph oric	**Bernie Sanders:** We are on **a race to the bottom**
(2) Interview with Nick Woodman (GoPro CEO) - Business and economics		Medium shots of Nick Woodman	Hand gesture, closed hand touching the temple and then suddenly	**Nick Woodman: The light bulb went off**

			opening up **Gesture function:** Representational / iconic	
(3) Interview with Sofia Coppola (Filmmaker) - Showbiz		Medium shot of Sofia Coppola	Hand and arm gesture, open hand, palm down. The hand and the arm suddenly move up **Gesture function:** Metaphoric/ representational	**Sofia Coppola:** In this kind of **over the top** situation

Table 4
Multimodal transcriptions of a selection of excerpts involving idiomatic expressions.

In example (1) in table (4) we may notice that the idiomatic expression used by Senator Bernie Sanders during an interview for the Sunday morning talk show ‘Meet the Press’ is “a race to the bottom”. This expression is tied to the socio-economic environment and indicates a state of competition where companies (or states or nations) attempt to undercut the competition’s prices by sacrificing standards, safety, regulations, wages and so on. As happens most of the time in TV interviews, the image is conveyed through medium shots, which frame the whole subject from the knees (or waist) up, and thus they allow the viewers to detect gestures and movements. The analysis of gestures reveals that the senator, while using the idiomatic expression, suddenly moves his hand and arms from up to down until he reaches the table. This gesture metaphorically recalls the idiomatic meaning of the expression, being the representation of something that falls, just like the economic situation, according to the interviewee. The second example represents the field of business and economics and (2) is taken from an interview with GoPro (an American technology company) CEO, Nick Woodman, aired on Charlie Rose, at the time one of the biggest interview talk shows in the USA. The idiomatic expression used by the interviewee is “the light bulb went off”, which is an informal American idiomatic expression meaning “to have a sudden realization or recognition”. Regarding gestures, we can appreciate in the two medium shots that the speaker lifts his hand to his head and opens it up so as to symbolize something that bursts. Therefore, the gesture, again, unveils the idiomatic meaning of the expression, which implies that something “turned on”. The last example in table (3) comes from the same talk show as example (2), and features an interview with the showbiz personality Sofia Coppola, a renowned filmmaker. While talking about one of her last movies, she defines the situation resorting to the idiomatic expression "over the top". The sense of such an expression is rather hyperbolic, meaning 'beyond normal', 'excessive'. In evaluating the gesture that co-occurs with the verbal message, it can be seen that the auteur raises her open hand and arm, metaphorically representing the idea of something that is well above the standard.

In table (5) below the same transcription was carried out for a selection of phrasal verbs co-occurring with gestures.

Interview title and domain	**a) Shot**	**b) Shot description**	**c) Gestures**	**d) Spoken language**
(4) Interview with US President Donald Trump - Political science		Long shot of President Donald Trump and TV anchor David Muir	Hand gesture, closed hand, index finger pointing, moving from right to left. Arm moving with the hand **Gesture function:** Deictic-indexical	**David Muir:** So, they are **paying us back**
(5) Interview with Meg Whitman (HP CEO) - Business and economics	 	Medium shots of Meg Whitman	Hand gesture, open hands, palms down. The hands interlace and the fingers are towards the palms **Gesture function:** Iconic-representational	**Meg Whitman: Put** old technology **together**
(6) Interview with Sofia Coppola (Filmmaker) - Showbiz		Medium shots of Sofia Coppola	Hand gesture, open hand, palm down and then up down. The hand fluctuates from right	**Sofia Coppola:** How to **fit in**

			and left. Imitating a snake	
			Gesture function: Metaphoric-representational	

Table 5
Multimodal transcriptions of a selection of excerpts involving phrasal verbs.

In example (4) there is one long shot (i.e. framing the whole figures of both interviewer and interviewee) of U.S. president Donald Trump and anchor-person David Muir. The attention here is on the interviewer himself who states that, according to Trump, Mexicans are “paying [us] back” for building the notorious wall. By uttering this phrasal verb, he makes a gesture with his right hand, using his index finger while moving from right to left. Such a deictic gesture has a clear indexical function, inasmuch as it positions and reinforces the transaction from ‘they’ to ‘us’. Example (5) is part of a business interview with Meg Whitman, the CEO of the Silicon Valley colossus Hewlett Packard. In the interview, she uses a series of phrasal verbs, such as “lever up [the balance sheet], “take out [the costs]”, “put [old technology] together”, to mention just a few. The shots show the interviewee’s fingers weaving together she utters the phrasal verb. The iconic nature of the gesture is crystal clear, concretely representing the idea of connection. The last example includes (6), which is another extract from Sofia Coppola’s interview. In the two medium shots, she uses the phrasal verb “fit in” and, at the same time, she moves her arm and hand miming a slithering snake, i.e. something that slips in. Such a representation, again, is metaphorically related to the non-compositional meaning of the phrasal verb, which, in this case, is to try to be accepted within a group.

5. Concluding remarks

Interviews broadcast on television are instruments of knowledge dissemination used to discuss a vast array of topics in a way that is accessible to those who turn on their television at home and tune to the network airing the show. They are, thus, crafted with the aim of encouraging the consumer not to change the channel, and to keep watching and enjoying the interview. The linguistic register used by the interactants is undoubtedly one of the elements that adapts the most to the need to reach a balance between clarity/specialization of information and entertainment that is dictated by both the TV and the dialogic format. The result is a product that is a hybrid mixture between conversational features, as well as specialized and planned discourse traits.

This research has proposed a pilot study of the pervasiveness of some orality indicators, i.e. idiomatic expressions and phrasal verbs, in political, business and showbiz interviews. A first quantitative analysis testing the occurrence of these linguistic phenomena showed that phrasal verbs and idiomatic expressions are more recurrent in business and political interviews, i.e. in more deliberately specialized interviews as compared to showbiz interviews (idiomatic expressions were more frequent in business interviews and phrasal verbs in political interviews). This result seems to point to the fact that, the more specialized the knowledge domain, the more language, when it is TV-mediated, is permeated with these involving and engaging expressions, perhaps in an attempt to smooth out the indirect interaction with the home audience and thus to build rapport with them.

Furthermore, using phrasal verbs and idiomatic expressions to a greater extent in political and business communication could not only be a stylistic choice making language more colourful and involving, but it could also have a simplifying function by, for example, substituting more complex words (e.g. Greco-Latinate words) or by just explaining very specialized and sophisticated concepts with figurative phrases, confirming the trend highlighted in Laudisio (2015) when studying specialized TV series. Even though more in-depth research would be needed to make precise claims, the overall tendency seems to be that formality and specialization tend to be mixed with colloquialism, so as to succeed in reaching the audience effectively with a more informal, direct and involving style.

Regarding the multimodal analysis of a selection of examples featuring these linguistic phenomena, it emerged that gestures are the nonverbal elements that are most often intertwined with the non-literal meaning of phrasal verbs and idiomatic expressions. Speakers in political and business interviews, in particular, employed gestures in synchrony with phrasal verbs and idiomatic expressions more often than in showbiz interviews. More interesting is the fact that idiomatic expressions most often co-occur with metaphoric gestures, i.e. representational movements that match and reinforce the idiomatic meaning expressed through the verbal. As for phrasal verbs, they are associated with beats (i.e. discourse organizing gestures that do not directly reflect the non-compositional meaning of the verb) to a greater extent in political interviews and showbiz interviews. In business interviews, instead, iconic gestures (i.e. concrete representations of the meaning of the phrasal verb) are the most recurrent typology.

In a nutshell, this preliminary study seems to confirm Biber and Reppen's (2002) claim that the frequency of idiomatic expressions and phrasal verbs is register and domain specific, namely their frequency varies according to the formality of the situation and to the knowledge domain we are talking about, adding that gesture use also seems to match with that in quite a consistent way. In fact, political and business interviews feature both more idiomatic expressions and phrasal verbs and co-occurring gestures than showbiz interviews, where perhaps other strategies, such as the use of humour, substitute the involving and social lubricant function covered by phrasal verbs and idiomatic expressions.

Bionote: Gianmarco Vignozzi holds a Ph.D. in English linguistics from the University of Pisa, where he teaches English language and linguistics and he carries out his research. His main interests lie in the area of ESP and corpus linguistics, ranging from TV-mediated political discourse (topic of his Ph.D. thesis) and filmic discourse analysis, to multimodal analysis and audiovisual translation. He has published several articles on these topics, both in journals and in edited volumes. His latest research works focus on the representation of spoken medical discourse in TV medical dramas and on the potentiality of multimodal stimuli to teach figurative expressions to foreign learners of English.

Author's address: gianmarco.vignozzi@fileli.unipi.it

Acknowledgements: This paper has been financed by the Italian Ministry for instruction (PRIN 2015 no. 2015TJ8ZAS)

References

Aijmer K. and Rühlemann C. 2014, *Corpus Pragmatics,* Cambridge University Press, Cambridge.

Baicchi A. and Rosca A. 2016, *Digging up the frequency of phrasal verbs in English for the police: the case of up,* in "CLAC" 67, pp. 273-296.

Baños R., Bruti S. and Zanotti S. (eds.) 2013, *Corpus linguistics and Audiovisual Translation: in Search of an Integrated Approach,* Special Issue of "Perspectives" 21.

Bednarek M. and Caple H. 2017, *The discourse of news values*, Oxford University Press, Oxford.

Biber D. and Reppen R. 2002, *What does frequency have to do with grammar teaching?*, in "SSLA" 24, pp. 199-208.

Biber D., Johansson S., Leech G., Conrad S. and Finegan E. 1999, *The Longman Grammar of Spoken and Written English*, Longman, London/New York.

Brants K. and Neijens P. 1998, *The infotainment of politics,* in "Political Communication" 15, pp. 315-355.

Bruti S. 2015, *Teaching the Use of Pragmatic Routines through Audiovisual Material*, in Crawford Camiciottoli, B. and Fortanet-Gómez, I. (eds.), *Multimodal Analysis in Academic Settings. From Research to Teaching,* Routledge, London/New York, pp. 213-236.

Bruti S. 2016, *Teaching Compliments and Insults in the EFL Classroom through Film Clips*, in Bianchi F. and Gesuato S. (eds.), *Pragmatic Issues in Specialized Communication*, Brill, Leiden, pp. 149-170.

Celce-Murcia M. and Larsen-Freeman D. 1999, *The grammar book: An ESL/EFL teacher's course*, Heinle and Heinle, Boston.

Clayman S. and Heritage J. 2002, *The news interview: Journalists and public figures on the air*, Cambridge University Press, Cambridge.

Cooper T. 1998, *Teaching idioms in English*, in "Foreign Language Annals" 2, pp. 255-266.

Crawford Camiciottoli B. 2007, *The Language of Business Studies Lectures*, Benjamins, Amsterdam.

Crawford Camiciottoli B. 2015, *Elaborating explanations during OpenCourseWare humanities lectures: The interplay of verbal and nonverbal strategies*, in Crawford Camiciottoli B. and Fortanet-Gómez I. (eds.), *Multimodal Analysis in Academic Settings. From Research to Teaching*, Routledge, London/New York, pp. 144-170.

Ekström M. 2001, *Politicians Interviewed on Television News*, in "Discourse & Society" 12 [5], pp. 563-584.

Erman B. and Warren B. 2000, *The idiom principle*, in "Text" 20 [1], pp. 29-62.

Fairclough N. 1998, *Political Discourse in the Media: An Analytical Framework*, in Bell, A. and Garrett, P. (eds.), *Approaches to Media Discourse*, Blackwell, Oxford.

Fairclough N. 2000, *New Labour, New Language?*, Routledge, London.

Freddi M. 2013, *Linguistica dei corpora*, Carocci, Bologna.

Furkó P. and Abuczki A. 2014, *English discourse markers in mediatised political interviews*, in "Brno Studies in English" 40 [1], pp. 45-64.

Gardner D. and Davies M. 2007, *Pointing Out Frequent Phrasal Verbs: A Corpus-Based Analysis*, in "TESOL Quarterly" 41 [2], pp. 339-359.

Heritage J. 1985, *Analyzing News Interviews: Aspects of the Production of Talk for an Overhearing Audience*, in van Dijk T.A. (ed.), *The Handbook of Discourse Analysis*, Academic Press, New York, pp. 95-119.

Jackendoff R. 1995, *The boundaries of the lexicon*, in Everaert M. *et al.* (eds.), *Idioms: Structural and Psychological Perspectives*, Hillsdale, Erlbaum, pp. 133-166.
Kendon A. 1980, *Gesture and speech: Two aspects of the process of utterance*, in Key R. (ed.), *Nonverbal communication and language*, Mouton de Gruyter, The Hague, pp. 207-227.
Kendon A. 2004, *Gestures: Visible Action as Utterance*, Cambridge University Press, Cambridge.
Kress G. 2010, *Multimodality. A Social Semiotic Approach to Contemporary Communication,* Routledge, London.
Laudisio A. 2015, *Popularization of Medical and Legal Language in TV Series*, in Bhatia V.K. *et al*. (eds.), *Variations in Specialized Genres*, Narr Francke, Tubingen, pp. 183-196.
Lauerbach G. 2006, *Discourse representation in political interviews: The construction of identities and relations through voicing and ventriloquizing*, in "Journal of Pragmatics" 38 [2], pp. 196- 215.
Liu D. 2003, *The most frequently used spoken American English idioms: A corpus analysis and Its implications*, in "TESOL Quarterly", 37 [4], pp.671-700.
Masi S. 2016, *Gestures in Motion in TED Talks: Towards Multimodal Literacy*, in Bonsignori V. and Crawford Camiciottoli B. (eds.), *Multimodality across Communicative Settings, Discourse Domains and Genres*, Cambridge Scholars, Newcastle-Upon-Tyne, pp. 146- 165.
McNeill D. 1992, *Hand and Mind: What the Hands Reveal about Thought*, The University of Chicago Press, Chicago.
McNeill D. 2005, *Gesture and Thought*, Chicago University Press, Chicago.
Milizia D. 2013, *Phrasal Verbs and Phrasal Units: Political Corpora within the Walls of the Classrom*, in Heller D., Desoutter C. and Sala, M. (eds.), *Corpora in Specialized Communication*, Cerlies Series, Bergamo, pp. 135-164.
Montgomery M. 2007, *The Discourse of Broadcast News*, Routledge, London.
Nunberg G., Sag I.A. and Wasow T. 1994, *Idiom*, in "Language" 70 [3], pp. 491-538.
O'Halloran K.L. 2010, *The Semantic Hyperspace: Accumulating Mathematical Knowledge across Semiotic Resources and Modalities*, in Christie F. and Maton K. (eds.), *Disciplinarity: Functional Linguistic and Sociological Perspectives*, Continuum, London.
Philip G. 2007, *Idioms*, in Malmkjaer K. (ed.), *The Linguistics Encyclopedia,* Routledge, London/New York.
Quirk R., Sidney G., Leech G. and Svartvik J. 1985, *A comprehensive grammar of the English Language*, Longman, London/New York.
Searle J.R. 1979, *Expression and Meaning: Studies in the Theory of Speech Acts*, Cambridge University Press, Cambridge.
Sinclair J. 1991, *Corpus, Concordance and Collocation*, Oxford University Press, Oxford.
Sinclair J. 1996, *The search for units of meaning*, in "Textus" 9 [1], pp. 75-106.
Sinclair J. 2008, *The phrase, the whole phrase and nothing but the phrase*, in Granger, S. Meunier F. (eds.), *Phraseology: An interdisciplinary perspective*, Benjamins, Amsterdam, pp. 407-410.
Sindoni M.G. 2013, *Spoken and Written Discourse in Online Interactions. A Multimodal Approach*, Routledge, London/New York.
van Dijk T.A. 1997, *What is Political Discourse Analysis*?, John Benjamins, Amsterdam.
Vignozzi G. 2016, *How idiomatic are Disney animated movies? The multimodal meaning-making of idioms*, in Bonsignori, V. and Crawford Camiciottoli, B. (eds.),

Multimodality across Communicative Settings, Discourse Domains and Genres, Cambridge Scholars, Newcastle-Upon-Tyne, pp. 236-264.

Weinberg A., Fukawa C. and Wiesner E. 2013, *Instructor gestures in proof-based mathematics lectures*, in Martinez M. and Castro A. (eds.), *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, University of Illinois at Chicago Superfine, Chicago.

Wildfeuer J. 2013, *Film Discourse Interpretation: Towards a New Paradigm for Multimodal Film Analysis*, Routledge, New York/London.

Wood D. 2010, *Formulaic language and second language speech fluency*: *Background, evidence, and classroom applications*, Continuum, London/New York.

Wray A. 2002, *Formulaic language and the lexicon*, Cambridge University Press, Cambridge.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 407-424
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p407

OLD WINE IN NEW BOTTLES
The case of the adjacency-pair framework revisited

FRANCESCA COCCETTA
UNIVERSITÀ CA' FOSCARI VENEZIA

Abstract – This study reports on the preliminary stages of a project concerned with presenting a systematic framework for annotating and analysing film dialogue which blends corpus annotation with adjacency pair theory (Schegloff, Sacks 1973). The adjacency pair model, which consists of closed sets of pairs, is applied to a searchable online corpus in which all the episodes of the U.S. medical drama *House M.D.* are transcribed and divided into scenes, the basic search unit. The article describes the modifications and adjustments to the original adjacency-pair model in order to accommodate the highly specialized dialogue of medical settings not envisaged by Schegloff and Sacks (1973) and the different interpersonal relationships between the characters, as well as the complex intersemiotic interactions characterizing TV film series.

Keywords: adjacency pairs; corpus annotation at discourse level; TV drama discourse.

1. Introduction

When Newton stated in 1675 "If I have seen further it is by standing on the shoulders of Giants" he wanted to highlight the fact that his discoveries were built on discoveries previously made by researchers, such as Kepler, and that the need to revisit, reconsider and adapt theoretical frameworks is an essential step in promoting advances in scientific research. The present article attempts to follow suit drawing on research into conversation analysis carried out some fifty years ago by Sacks and Schegloff (Schegloff 1968, 1980; Schegloff, Sacks 1973; Sacks *et al*. 1974) and in particular revisiting their adjacency-pair framework, considered by many as a limiting framework for the analysis of authentic conversation (e.g. Sinclair, Coulthard 1975; Eggins, Slade [1997] 2006). The research question addressed is whether the model, with adaptations and integrations, can be successfully applied to the scripted discourse of TV dramas, a kind of discourse for which this model was not conceived but which in the years since the model was first drawn up has become an object of considerable academic research and discussion (Taylor 1999; Rossi 2003;

Pavesi 2005; Quaglio 2009), at the very least on a par with the studies of spontaneous oral discourse that dominated the development of discourse models half a century ago (Labov, Waletzky [1967] 1997; Sinclair, Coulthard 1975; Burton 1978; Coulthard, Brazil 1979). Put another way, the article asks whether the affordances of the digital age, which potentially include the application of corpus-based approaches to digitalized TV dramas, are such as to transcend the model's perceived shortcomings by foregrounding and advocating its hidden potentialities.

The preliminary research presented in this article focuses on the scripted discourse of the U.S. medical drama *House M.D.* and is part of the *MWSWeb House Corpus* project (*http://mws.itd.cnr.it*), a searchable online corpus in which all the 177 episodes of the U.S. medical drama *House M.D.* are transcribed and divided into scenes, the basic search unit. In particular, this strand of investigation of the project will consider the following research questions:

1. how far can the adjacency-pair framework be applied to House M.D. scene closures?
2. how relevant is the concept of a multimodal adjacency pair in which the first part is enacted linguistically and the second part by a non-linguistic response (e.g. silence or a specific body movement) and viceversa?

The article is arranged as follows: *Section 2* reviews the literature on the adjacency-pair model and its application to authentic and scripted discourse. *Section 3* provides a description of how the adjacency-pair model can be turned into a search tool and of what modifications and adjustments are needed in the furtherance of this goal. *Section 4* discusses the hurdles of the *MWSWeb House Corpus* project regarding the analysis of adjacency pairs. Finally, *Section 5* draws some conclusions and possible future developments.

2. Literature review

One of the basic tenets of Conversation Analysis (henceforth CA), and an "analytical innovation" (Goodwin, Heritage 1990, p. 287) for that time, is the concept of interactional sequence. This led Schegloff and Sacks (1973) to develop the concept of adjacency pair, which was defined as the basic unit of conversational exchange. The adjacency pair consists of two adjacent utterances produced by different speakers and functionally related to each other as the second utterance is determined by the first utterance. For Schegloff and Sacks, adjacency pairs typically included *question/answer*, *greeting/greeting*, *complaint/denial*, *offer/accept*, *request/grant*, *compliment/response*, *challenge/rejection*, and *instruct/receipt*. There were, they argued, pairs such as *question/answer*, *greeting/greeting* and *compliment/response* which typically expect just one response. On the contrary, other sequences allowed the second speaker in conversation to choose from two options, that they described as a preferred one and a dispreferred one. A case in point is the first-pair part *offer*. The preferred second-pair part is *accept*, while the dispreferred is *refusal*. The latter is considered more face-threatening because it requires the speaker to draw on mitigating strategies to explain the reason for the refusal and ensure the cooperative principle in conversation (Grice 1975).

The adjacency-pair model is based on the property of "conditional relevance" (Schegloff 1972) whereby the second-pair part is conditioned by the first-pair part as is the case of the *question/answer* sequence. However, as Schegloff (1972) observed, there are cases where a question is not followed by an answer, but by another question, yet the answer is not seen as absent. Schegloff called these sequences "'insertion sequence' or 'inserted sequences,' because between an initial question and its answer there is inserted another question-answer sequence" (Schegloff 1972, p. 78). As an example Schegloff (1972, p. 78) provides the following sequence:

A: Are you coming tonight?
B: Can I bring a guest?
A: Sure.
B: I'll be there.

where the *question/answer* sequence *Can I bring a guest?/Sure* is an insertion sequence.

For many years, CA was applied only to the analysis of spontaneous conversation, or talk-in-interaction in CA's terms. However, Goodwin and Heritage (1990, p. 289) observe that "ordinary conversation is the point of departure for more specialized communicative contexts (e.g. the legal process, the educational system, the medical encounter), which may be analysed as embodying systematic variations from conversational procedures". Indeed, the adjacency-pair framework has been applied not only to everyday conversations, but also to various kinds of institutional encounters including medical consultations (West 1983; Frankel 1990), law courts (Atkinson, Drew 1979), news interviews (Heritage 1985; Heritage, Greatbatch 1991; Greatbatch 1988), classroom lessons (McHoul 1978; Mehan 1985), and psychotherapeutic interactions (Peräkylä *et al.* 2008).

However, research has also demonstrated the limit of the adjacency-pair framework by pointing out that the majority of conversational exchanges consist of more than two utterances. For example, Sinclair and Coulthard (1975) observed that the basic unit of analysis of classroom discourse is a three-part exchange consisting of the following three moves: 1) Initiation; 2) Response; and 3) Feedback. The Initiation is spoken by the teacher and followed by the student's Response; finally, the student's Response is followed by the teacher's Feedback, which aims to indicate whether the student's Response is right or wrong. This formula was then revised by Coulthard and

Brazil (1979) who proposed the following exchange structure, consisting of from a minimum of two moves up to a maximum of six moves: 1) Initiation; 2) (Response); 3) (Initiation); 4) Response; 5) (Feedback); and 6) (Feedback).[1] Similarly, although in passing, Goffmann (1976) noted the unsatisfactoriness of the adjacency-pair framework:

> [a] response will on occasion leave matters in a ritually unsatisfactory state, and a turn by the initial speaker will be required, encouraged, or at least allowed, resulting in a three-part interchange; or chains of adjacency pairs will occur (albeit typically with one, two, or three such couplets), the chain itself having a unitary, bounded character. (Goffman 1976, p. 272)

The result has been that other models have taken over. An example is Eggins and Slade's ([1997] 2006) model based on Halliday's (e.g. 1973, 1978) theory of language.

Although the adjacency-pair framework appears to have had its limitations in the description of authentic interactions, conversely, the framework would appear to have had some success with scripted texts such as dramas, as is the case of the TV drama *House M.D.* considered in this article. In this regard, Coulthard ([1977] 1985) states that "[d]rama texts, being scripted for the performing of pseudo-conversations, can be successfully approached with techniques originally developed to analyse real conversation" (Coulthard [1977] 1985, p. 182). Thus, over the years, while studies relating to the model's application to spontaneous oral discourse in English appear to have dwindled, on the contrary studies that apply the model to scripted discourse appear to have slowly gained ground (see Table 1 for a list of studies on scripted discourse analysed using CA) in spite of criticisms of the use of CA as a tool to investigate non-spontaneous interactions (Emmison 1993; Schegloff 1988).

Study	Genre	Aspect investigated
Coulthard [1977] 1985	*Othello*	question/answer sequence
Herman 1998	plays by Shakespeare, Pinter, Osborne, Beckett, and Shaw	systematics of turn-taking
Piazza 1999	plays by Haye, Norman, Shepard and Pinter	conversational repairs
Piazza 2006	Italian melodrama	confrontational talk
Stokoe 2008	TV series *Friends*	adjacency pairs
Chepinchikj, Thomposon 2016	three Woody Allen films	adjacency pairs

Table 1
Some studies on scripted discourse using CA.

Using CA, Coulthard ([1977] 1985) analyses the *question/answer* sequence in Shakespeare's *Othello* and observes that the tragedy is brought about by Iago's

[1] The moves in round brackets are optional.

avoidance of answering Othello's questions about Cassio. In so doing, Iago arouses suspicion of his wife's adultery with Cassio as it suggests to Othello that Iago is concealing something.

Herman (1998) explores the systematics of turn-taking in dramatic discourse, particularly Shakespeare's, Pinter's, Osborne's, Beckett's, and Shaw's, and observes that in their plays dramatists make use of varied sequencing options (e.g. adjacency pairs, side sequences, remedial sequences) and pattering which contribute to interaction management and the creation of situations. She states that:

> various interactive paths are possible. The normative rules are those that maximize on co-operation and harmony in talk, but the rules can be manipulated to create other situations – of conflict, isolation, separation – since these are also potential realities of inter-personal communication and talk. The paths taken and the effects achieved are also owing to the strategic use of the system. The variety and creativity of the uses of the turn-taking system are legion, since each segment in a play creates its own specificities of situation and character through the enacted patterns of participants' talk. (Herman 1998, p. 162)

Work on dramatic discourse using CA has been carried out by Piazza (1999), as well. In her study, she analyses the mechanisms for conversational repairs in four modern plays (Haye's *Skirmishes*, Norman's '*Night, Mother*, Shepard's *Fool for Love*, and Pinter's *The Caretaker*). Her study shows that, when compared to real-life exchanges, dramatic discourse makes use of deviant discourse procedures; the use of ordinary communicative strategies, on the other hand, takes place only for special dramatic purposes.

Piazza (2006) also analyses confrontational talk, defined as a break in conversational talk, in a corpus of Italian films belonging to the melodrama genre, namely narratives characterized by dramatic passion. In particular, she focuses on the *question/answer* adjacency pair and the strategies used in the answer slot to deal with aggressive questions or questions perceived as aggressive. Piazza (2006) observes that the following three strategies are used: 1) reply for an answer, i.e. the answer does not satisfy the original question; 2) silence as an answer, i.e. the answer consists of a silence; and 3) questions as replies, i.e. the answer is a new question.

CA has also been applied to the language of sitcoms and films as demonstrated by the studies by Stokoe (2008) and Chepinchikj and Thomposon (2016). Stokoe (2008) uses CA to examine the production of humour in the American sitcom *Friends* and observes that the breaching of conversational "rules", particularly the preference organization in adjacency pairs, and interruptions in the structure of turn-taking are used on purpose to elicit laughter from the TV audience. Chepinchikj and Thompson (2016) apply CA to segments taken from three Woody Allen films (*Husbands and Wives*, *Melinda and Melinda*, and *You will Meet a Tall Dark Stranger*). In some scenes, they observe a lack of adjacency pairs which reveals a disconnection between the characters portrayed; on the other hand, when present, the adjacency pairs tend to completely lack the second-pair part in *question/answer* and *request/answer* sequences, or this is delayed, i.e. the second-pair part does not immediately follow the first-pair part but occurs later in the dialogue. They conclude that "[t]hese marked types of adjacency pairs are used as an interactional device to evade and deflect the course of conversation to the benefit of the character who is performing that action" (Chepinchikj, Thompson 2016, p. 48).

What emerges from these studies is a skilful use of some of the features of conversational exchanges with a recurrent deviation from the conversational norms of turn-taking. It can be thus expected that a similar pattern will be found in *House M.D.*, too. In passing, we may recall that Dr. House, the main character of the series, is a rather special doctor who accepts only puzzling cases and does whatever it takes to solve them, including flouting hospital rules and opting for ethically-questionable procedures. For this reason, he frequently comes into conflict not only with his boss, Dr. Lisa Cuddy, the hospital administrator and Dean of Medicine, but also with his good friend, Dr. James Wilson, and his team.

3. Analysis hypothesis and sampling

3.1. Adjacency pairs as a search tool

To date, to the author's knowledge, the long-standing tradition of the adjacency-pair framework's application to scripted discourse has not included the possibility of using the model as a way of searching through corpora. Traditionally, corpus studies are based on word or collocation patterns as well as multi-word units as demonstrated by the extensive literature triggered by Sinclair's (1991) seminal work *Corpus, Concordance, Collocation*. As regards research into spoken corpora, corpus searching based on larger units, such as complete utterances, has rarely been taken into account, as there would probably only be one or two examples, or no examples of patterned use. One exception is the attempt to explore larger units made by Adolphs (2008) who uses a corpus-based approach to analyse the pragmatic function of utterances in the Cambridge and Nottingham Corpus of Discourse in English (CANCODE) (McCarthy 1998). Adolphs points out the difficulty of combining a corpus-based approach to analyse language functions as "the pragmatic approach starts from a functional perspective, while the corpus approach starts from a lexicogrammatical perspective" (Adolphs 2008, p. 44). In this regard, Swales states:

> starting with a lexical item […] is relatively straightforward, at least initially, because one can be fairly sure of capturing all the tokens […]. Starting with a functional category, in contrast, means searching the grammatical and pragmatic literature as well as racking one's brain in order to come up with a list of possible realizations. (Swales in Adolphs 2008, p. 9)

Swales' statement is demonstrated by Adolphs with some examples of the lexicogrammatical realizations of the *suggest* function. However, Adolphs also shows that the concordance lines obtained need to be filtered to eliminate the "undesired" lines, i.e. those concordance lines where the lexicogrammatical realization does not realize the function under analysis. A case in point is the expression *why don't you* which, besides being predominantly used to put forward a suggestion, is also used with the meaning *for what reason*.

In the light of Adolphs' (2008) study and previous research on functional concordancing (Coccetta 2008, 2011), in this article it is argued that the abstract nature of the adjacency-pair model, on the other hand, could be the basis for a search model that overcomes the production of *one-offs*, i.e. just one or two examples, as well as the difficulties described by Adolphs (2008). Besides, as recalled above in *Section 2*, the literature on adjacency pairs demonstrates that patterns of use do emerge, for example, where a question is followed by a question rather than an answer, while when it comes to dramatic discourse, the model would point out patterns of use that deviate from what typically occurs in authentic discourse. As a consequence, it would be useful to have the possibility of searching an entire corpus to discover the nature and incidence of such pairs. For instance, it may be useful to see whether the follow-up question is conflictual in nature, or designed to seek clarification, or merely requiring information, possibly misheard, to be repeated.

However, such an undertaking is time-consuming so that strategies need to be developed to make the task more manageable. A pilot study is what is called for and what is reproduced in this article *vis-à-vis* the *House Corpus*. In particular, the study's final aim is creating an adjacency-pair search interface to be integrated in the *MWSWeb House Corpus* project (*http://mws.itd.cnr.it*). The project makes the *House Corpus* available online along with the possibility of accessing each scene individually (Baldry, in press; Taibi *et al.*, in press).

3.2 Adjacency pairs in the House Corpus

A first step in the construction of an adjacency-pair search interface is manual analysis that identifies the occurrence of adjacency pairs in *House M.D.* scene closures. In this preliminary stage of the project, it was decided to focus on scene closures for two reasons: one practical and one theoretically grounded. From a practical point of view, the detection of the last two lines of a scene is relatively simple thanks to the pre-existing scene division in the corpus (see Baldry 2016) that establishes the start and the end of a scene. From a theoretical point of view, it can be argued that a scene in a film script corresponds to what Schegloff and Sacks (1973, p. 289) refer to as "the unit 'a single conversation.'" Some terminological explanations are in order. In particular we need to define what a *scene* and a *single conversation* are.

McKee (1997, p. 21) states that the goal of a movie or TV series scriptwriter is "a good story *well told*" (italics in the original). The story has a structure consisting of the following units, arranged in ascending order of size (smaller to larger): *beat*, *scene*, *sequence*, and *act*. What is of interest for the present study is the scene, also known as story event. A scene is defined as "a meaningful change in the life situation of a character that is expressed and experienced in terms of a value" (McKee 1997, p. 33). By value, McKee means "the universal qualities of human experience that may shift from positive to negative, or negative to positive, from one moment to the next" (McKee 1997, p. 34). Examples of positive/negative values are love/hate, strength/weakness, and courage/cowardice. In other words, a scene is a turning point in the plot, "a story in miniature" (McKee 1997, p. 233).

In conversation analysis, a single conversation is a unit of analysis defined as a conversation activity where one speaker speaks at a time and speaker change recurs. As a result, a single conversation consists of a sequence of turns. Besides, it is structurally organized in segments such as openings, middles and closings.

Schegloff and Sacks (1973) paid particular attention to the problem of what they called the closings of conversations, specifically to how a conversation is brought to a close. They pointed out that adjacency pairs are typically used in terminal exchanges (i.e. the ending sequences of 'a single conversation'). It can be thus concluded that it is highly probable to find adjacency pairs at the end of the scenes. This hypothesis is supported by Herman's (1998) work on dramatic discourse. In particular, she observes that adjacency pairs "permit the closure of interactive business with high economy" (Herman 1998, p. 124).

Examples 1 and 2 support the hypothesis that the penultimate line often contains the first-pair part of the adjacency pair while the last contains the second-pair part. In the scene shown in Example 1, the main characters are Jonah, a four-year-old patient, his mother, Melinda, and Dr. Foreman and in the scene closure we can see the *question/answer* adjacency pair.

[Jonah starts giggling] Melinda: Does the pill tickle, sweetie? Foreman: He can't feel it. What's funny, your brother? [Foreman looks into Jonah's eyes with his light. Jonah continues to laugh] Melinda: What's happening? Foreman: I don't know.

Example 1
Scene closure containing a *question/answer* adjacency pair (from *House M.D.*, S5 E08).

In Example 2, Dr. Chase instructs Stewart, the patient, to count down from ten and he complies. This is an example of the *instruct/receipt* adjacency pair. In passing, we may recall that in the adjacency-pair model the term *receipt* means compliance.

Stewart: (grabbing Cameron's hand) I know I'm a pain in the ass. I know it would have been easy just to walk away, but… You're a good person. [Cameron looks like she is having doubts about the deception. Stewart obviously trusts her] [Chase administers the anesthesia] Chase: Count down from 10. Stewart: 10, 9, 8… [Foreman and Kutner come in and they all prepare to move Stewart to the hospital]

Example 2
Scene closure containing an *instruct/receipt* adjacency pair (from *House M.D.*, S5 E07).

Similar to face-to-face interactions, TV series rely on the interplay of different semiotic resources, including action. Thus, it can be assumed that one pair in the sequence can be realized non-linguistically with action. A case in point is Example 3.

Kutner: If we don't give him a new kidney, he's dead in a week. We have to transplant. House: When your remote has a dead battery, you don't replace it with another dead battery. Whatever's killing the dad's kidneys is gonna kill the kid's too. Taub: It is possible the sleep issues aren't medically related. Maybe they're both insomniacs due to the turnpike noise. Maybe they both drink too much coffee, maybe – Thirteen: (approaching) House is right. Taub: How do you know? Thirteen: Because the daughter's sweating blood. If they have the same thing, it means there has to be a common cause. Which means it has to be a toxin, infection, or genetic. And since we've ruled out infections and toxins – Kutner: It narrows it down to any one of a dozen genetic disorders, each of which takes more than a week to run. House: Call Foreman. Get to work. [The team leave]

Example 3
Scene closure containing an *instruct/receipt* adjacency pair (from *House M.D.*, S5 E06).

In this example, the first-pair part of the adjacency pair is realized linguistically, that is, "Call Foreman. Get to work", while the second-pair part is realized non-linguistically: House's diagnostic team (Kutner, Taub and Thirteen) leave in order to comply with and carry out their boss's instruction. Here, we do not see the outcome, but it is highly probable that the team will comply because of the interpersonal relations between themselves and House, particularly the professional hierarchy that characterizes this relation (House is the boss, the team are the subordinates in the pecking order – indeed they are known affectionately by fans as House's ducklings).

In this respect, a preliminary analysis of Season 5 reveals a recurrent pattern, that is, House's instructions to his team tend to be linguistic while the compliance is non-linguistic. In other words, as the Series' logic proverbially dictates, House always gets the last word. Indeed, in this case and many others, the first-pair part of the adjacency pair is not realized in the penultimate line, as previously hypothesized, but in the last line. Thus, we can find sequences where the last line contains the first-pair part of the adjacency pair, while the second-pair part is realized visually in the same scene, or in a following one, with a gesture or other actions.

We can thus posit that in the *House Corpus* two types of adjacency pair exist: Adjacency Pair Type 1, where the first-pair part and the second-pair part of the adjacency pair are respectively realized in the penultimate and last lines of the scene script, and Adjacency Pair Type 2, where the first-pair part is realized just in the last line of the scene script and the second-pair part in a "missing line".

3.3. Incorporating the adjacency-pair model into the House Corpus

On the basis of the above pilot study, a provisional model has been incorporated into the *House Corpus* to annotate each scene for Adjacency Pair Type 1 and Adjacency Pair Type 2. This is illustrated in Figure 1.

Figure 1 shows that the scene under analysis is provided with the transcript and some annotation options which include: 1) Location type (e.g. Hospital or Elsewhere); 2) Event type (e.g. Case Discussion, Public or Private); and 3) Interaction type. What is of interest for this study is the last category. The drop-down menu next to Interaction type enables the annotator to choose from Adjacency Pair Type 1 and Adjacency Pair Type 2. If Adjacency Pair Type 1 is selected (see Figure 2), the annotator is provided with two drop-down menus, one for each line in the dialogue, that enable them to annotate the penultimate line and the last line for the first-pair-part options (e.g. greeting, question, and instruct) and the second-pair-part options (e.g. greeting, answer, and receipt) respectively.

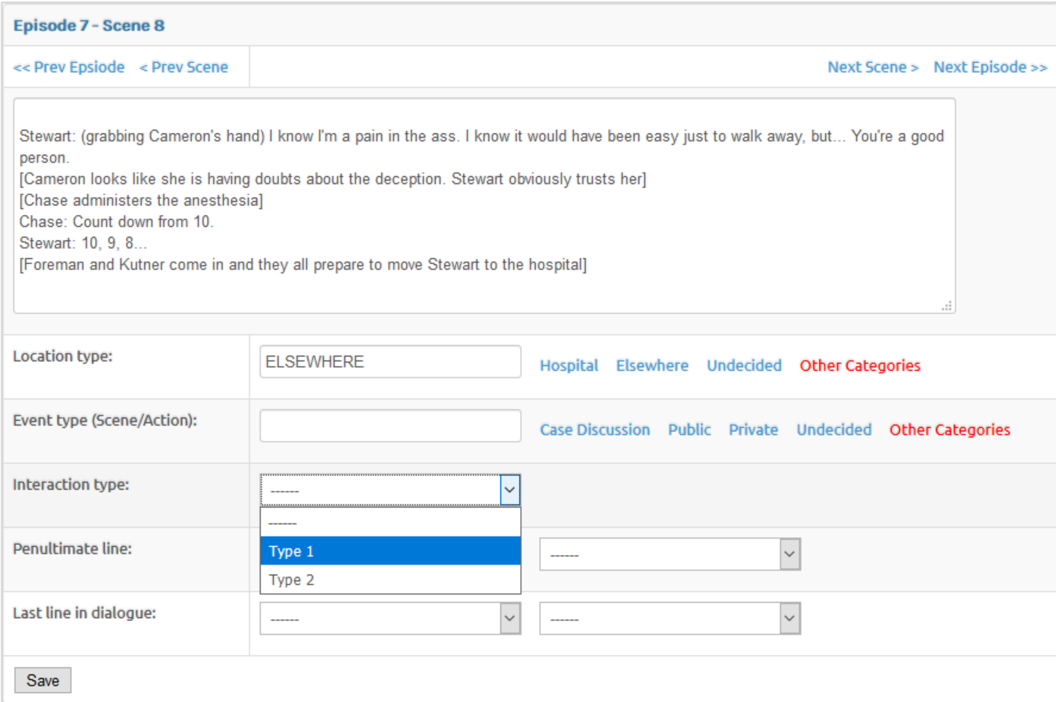


Figure 1
Interface for the annotation of the *House Corpus*.

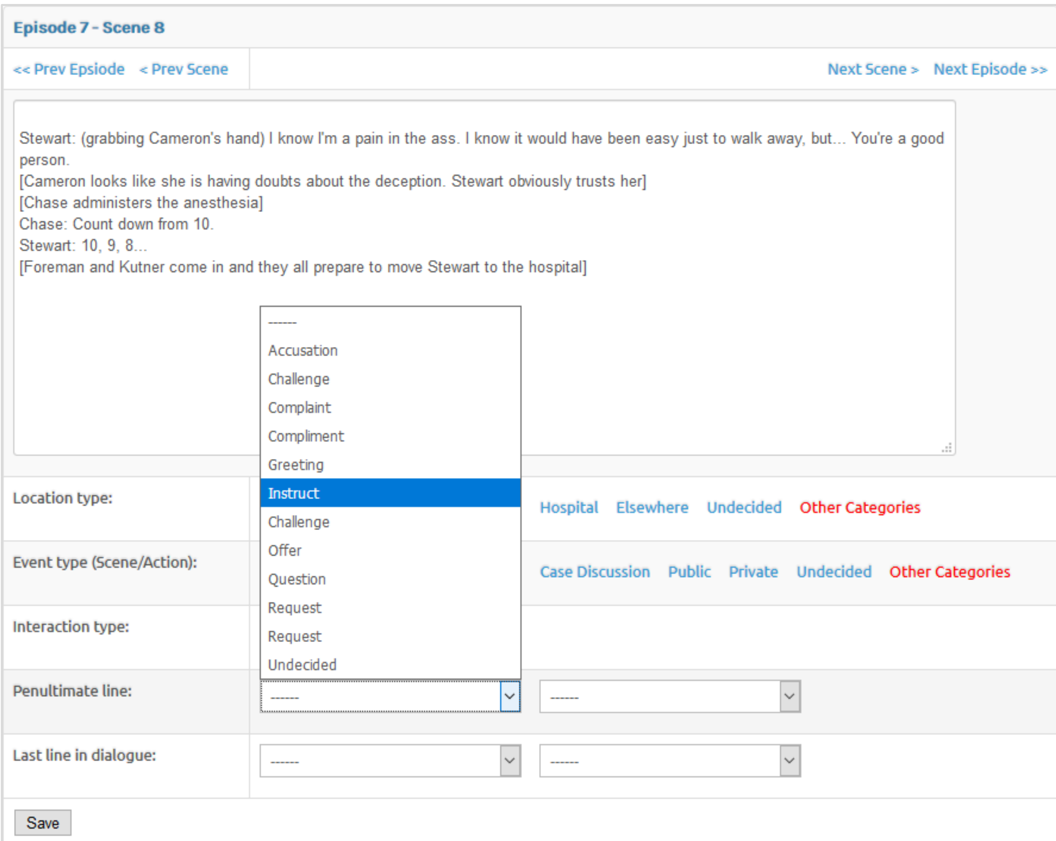


Figure 2
Interface for the annotation of Adjacency Pair Type 1 in the *House Corpus*.

If Type 2 is selected (Figure 3), the annotator is provided with a drop-down menu for the last line in the dialogue which enables them to annotate it for the first-pair-part options (e.g. greeting, question, and instruct) and a drop-down menu for the missing line in the dialogue which enables them to decide whether the second-pair part is realized non-linguistically (e.g. with an action) or not yet known, i.e. as in the case of the realization in a following scene.

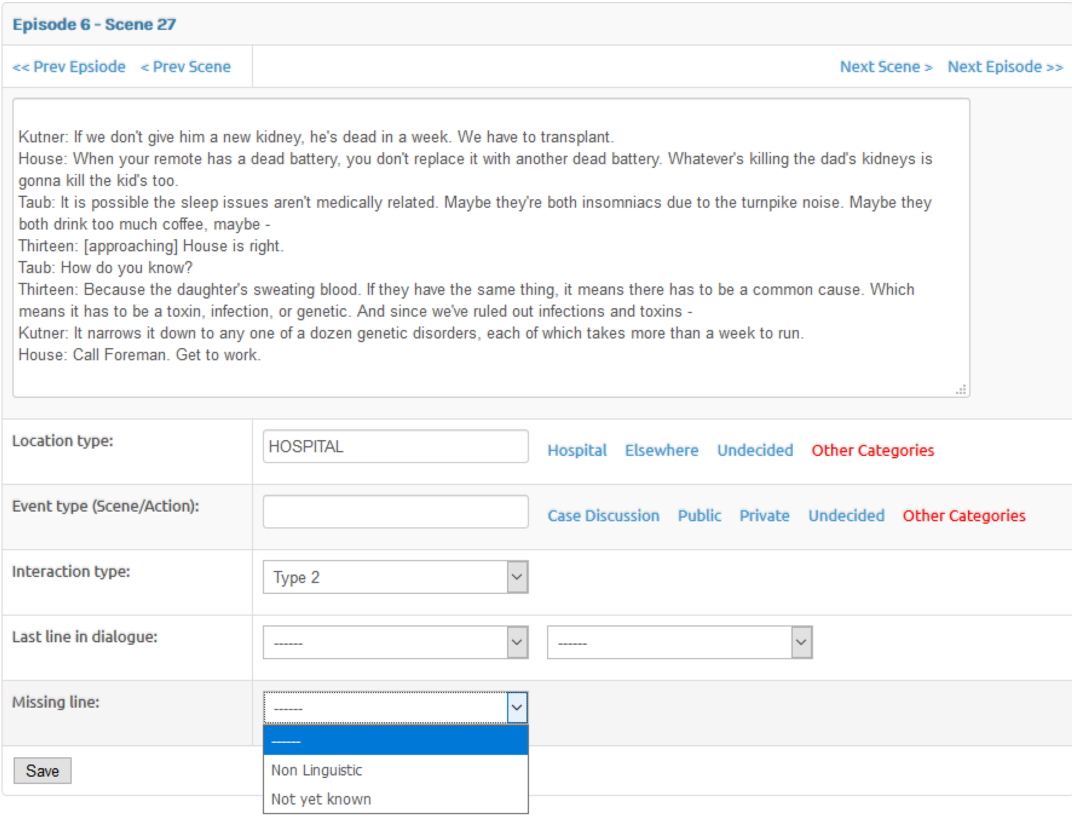


Figure 3
Interface for the annotation of Adjacency Pair Type 2 in the *House Corpus*.

In addition, besides annotating each line for the type of first- and second-pair part, the annotator can specify whether this is realized linguistically or non-linguistically. The drop-down menus provide the annotator with the option "undecided" to be selected when the type of adjacency pair is not clear. This constitutes a second step in a strategy designed to make adjacency-pair searching a reality. As already pointed out above, this has been made possible by the pre-existing division of the corpus into scenes. Thus, an annotator can tag the presence of an adjacency pair in a scene and a user can check out the results of such annotation efforts using the search model, as Figure 4 shows in a simplified way.

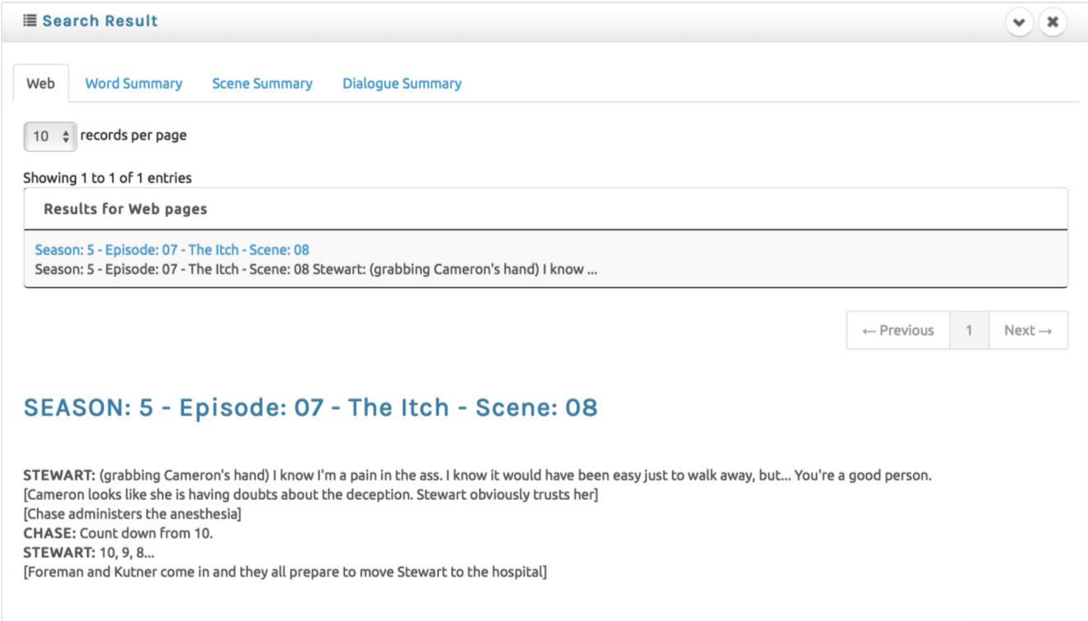


Figure 4
Search result for Adjacency Pair Type 2 in the *House Corpus*.

4. Discussion

Although in recent years annotation at discourse level has made some progress as shown by Archer *et al.* (2008) *vis-à-vis* pragmatic annotation, it is still a high-risk business due to the apparent shortage of annotation tools. The pilot study sketched above has considered the adjacency-pair framework (Schegloff, Sacks 1973) and has shown that the model can be incorporated into an interface to annotate scripted discourse, specifically that of the TV series *House M.D.*, and search for the various types of adjacency pairs, thus making adjacency pair searching a viable proposition, one capable of producing useful discourse patterns as the *instruct/repair* adjacency pair shown in Example 3 above where the *instruct* part, addressed to House's team, is linguistically realized by House himself while the *repair* part is realized non-linguistically by his team who leave, an indication of their acceptance. In addition, in the light of the multimodal nature of the TV series, the study has pointed out the need to modify and adjust the framework in order to ensure that the non-linguistic realizations of one of the two pair parts can be easily and successfully annotated.

There are, of course, many hurdles to be overcome. First, only scene closures are currently envisaged. However, there is a need to find ways to extend the model to other parts of the scenes to better understand the mechanisms for interaction among the various characters and shed some light on their interpersonal relations. In this regard, we can recall Goodwin and Heritage's observation that "CA's focus on conversation between acquaintances, friends, and siblings offers an opportunity to determine what is distinctive about interactions involving asymmetries as status, gender, and ethnicity" (Goodwin, Heritage 1990, p. 289). One of these ways includes the use of semi-automatic annotation. For example, one of the functionalities under discussion is the automatic identification of question types owing to the presence of the "?" character. This facilitates and speeds up annotation of the *question/question* adjacency pair discussed above.

Second, the examples provided in *Section 3.3* have illustrated the fact that when analysing the scene closures of *House M.D.* two adjacency pair types need to be posited. Besides what was defined as Adjacency Pair Type 1, where the first-pair part and the second-pair part of the adjacency pair are respectively realized in the penultimate and last lines in dialogue, an Adjacency Pair Type 2 also needs to be included, where the first-pair part is in the last line and the second-pair part is in a missing line. There is, however, some evidence to suggest the predictability of these types, for example, in relation to non-verbal responses in final adjacency pairs in Season 5 where a recurrent plotline is House's hostile relationship with his new recruits. Such fine-grained analysis goes beyond the scope of this first-stage "feasibility" report but suggests the need for more exhaustive accounts of this and other TV series as well as further research into corpus-based approaches to discourse studies as a way of simulating the alternative possibilities and strategies available to interactants at specific points in the construction of discourse, of which closure, however significant, is just *one* aspect.

5. Conclusions

The present article has considered the adjacency-pair model developed by Schegloff and Sacks (1973) from different perspectives. First of all, it has reviewed some of the literature on the adjacency pair and its applications to the analysis of authentic conversations (or talk-in-interactions in CA's terms) as well as institutional encounters and to scripted discourse, particularly plays, television series and movies. The article has also pointed out the lack of corpus-based studies where the model has been applied. In the case of the studies on scripted discourse mentioned above, it can be argued that the availability of corpus tools and techniques for CA analysis might well have facilitated the researcher's work. Secondly, the article has described the application of the adjacency pair model to the American TV drama *House M.D.* and demonstrated the need to modify and adjust the model in order to accommodate the highly specialized dialogue of medical settings not envisaged by Schegloff and Sacks (1973) and the different interpersonal relationships between the characters, as well as the complex intersemiotic interactions characterizing TV film series. In so doing, the study has shown the possibility of applying a model based on adjacency pairs to corpus studies thus going beyond the word level which, traditionally, has received much attention in the development of annotating techniques (e.g. POS tagging, lemmatization, semantic annotation). Further steps in the project, some of them briefly suggested above, will hopefully allow greater insights to be made that will be reported in subsequent publications.

Bionote: Francesca Coccetta is a tenured Assistant Professor at Ca' Foscari University of Venice. She holds a doctorate in English Linguistics from Padua University where she specialized in multimodal corpus linguistics. Her research interests include multimodal discourse analysis, learner corpus research, and the use of e-learning in language teaching and learning. She has published several articles on these topics in specialized journals and in various books. Her publications include the monograph *Access to Discourse in English through Text Analysis: A Preparatory Guide for Undergraduate Students*, Ibis, Pavia, 2016.

Author's address: francesca.coccetta@unive.it

References

Adolphs S. 2008, *Corpus and Context. Investigating Pragmatic Function in Spoken Discourse*, Benjamins, Amsterdam/Philadelphia.
Archer D., Culpeper J. and Davies M. 2008, *Pragmatic Annotation*, in Lüdeling A. and Kytö M. (eds.), *Corpus Linguistics. An International Handbook. Volume 1*, Walter de Gruyter, Berlin/New York, pp. 613-642.
Atkinson J.M. and Drew P. 1979, *Order in Court: The Organization of Verbal Interaction in Judicial Settings*, Macmillan, London.
Baldry A. 2016, *Multisemiotic Transcriptions as Film Referencing Systems*, in "InTRAlinea: Online Translation Journal". http://www.intralinea.org/specials/article/2195 (30.04.2018).
Baldry A. in press, *Introduction*, in Baldry A., Loiacono A. and Bianchi F. (eds.), *Representing and Redefining Specialised Knowledge: Medical Discourse*, ESE – Salento University Publishing.
Burton D. 1978, *Towards and Analysis of Casual Conversation*, in "Nottingham Linguistics Circular" 17 [2], pp. 131-159.
Chepinchikj N. and Thompson C. 2016, *Analysing Cinematic Discourse Using Conversation Analysis*, in "Discourse, Context and Media" 14, pp. 40-53.
Coccetta F. 2008, *First Steps towards Multimodal Functional Concordancing*, in "Hermes" 41, pp. 43-58.
Coccetta F. 2011, *Multimodal Functional-Notional Concordancing*, in Frankenberg-Garcia A., Flowerdew L. and Aston G. (eds.), *New Trends in Corpora and Language Learning*, Continuum, London/New York, pp. 121-138.
Coulthard M. [1977] 1985, *An Introduction to Discourse Analysis. Second Edition*, Routledge, London/New York.
Coulthard M. and Brazil D. 1979, *Exchange Structure*, Discourse Analysis Monograph No. 5, English Language Research, University of Birmingham, Birmingham.
Eggins S. and Slade D. [1997] 2006, *Analysing Casual Conversation*, Equinox, London/Oakville.
Emmison M. 1993, *On the Analyzability of Conversational Fabrication: A Conceptual Inquiry and Single Case Example*, in "Australian Review of Applied Linguistics" 16, pp. 83-108.
Frankel R.M. 1990, *Talking in Interviews: A Dispreference for Patient-Initiated Questions in Physician-Patient Encounters*, in Psathas G. (ed.), *Interaction Competence*, University Press of America, Lanham, MD, pp. 231-262.
Greatbatch D. 1988, *A Turn-Taking System for British News Interviews*, in "Language in Society" 17 [3], pp. 401-430.
Goffman E. 1976, *Replies and Responses*, in "Language in Society" 5, pp. 257-313.
Goodwin C. and Heritage J. 1990, *Conversation Analysis*, in "Annual Review of Anthropology" 1, pp. 283-307.
Grice H.P. 1975, *Logic and Conversation*, in Cole P. and Morgan J.L. (eds.), *Syntax and Semantics: Speech Acts. Volume 3*, Academic Press, New York, pp. 41-58.
Halliday M.A.K. 1973, *Explorations in the Functions of Language*, Edward Arnold, London.
Halliday M.A.K. 1978, *Language as Social Semiotic*, Edward Arnold, London.
Heritage J. 1985, *Analyzing News Interviews: Aspects of Talk for an Overhearing Audience*, in Dijk T.A. (ed.), *Handbook of Discourse Analysis Volume 3*, Academic, New York, pp. 95-119.

Heritage J. and Greatbatch D. 1991, *On the Institutional Character of Institutional Talk: The Case of News Interviews*, in Boden D. and Zimmerman D.H. (eds.), *Talk and Social Structure*, University of California Press, Berkeley, pp. 93-137.
Herman V. 1998, *Dramatic Discourse. Dialogue as Interaction in Plays*, Routledge, London/New York.
Labov W. and Waletzky J. [1967] 1997, *Narrative Analysis: Oral Versions of Personal Experience*, in "Journal of Narrative and Life History" 7 [1-4], pp. 3-38.
McCarthy M. 1998, *Spoken Language and Applied Linguistics*, Cambridge University Press, Cambridge.
McHoul A. 1978, *The Organization of Turns at Formal Talk in the Classroom*, in "Language in Society" 7, pp. 183-213.
McKee R., 1997, *Story. Substance, Structure, Style and the Principle of Screenwriting*, Harper-Collins, New York.
Mehan H. 1985, *The Structure of Classroom Discourse*, in Dijk T.A. (ed.), *Handbook of Discourse Analysis Volume 3*, Academic, New York, pp. 120-131.
Pavesi M. 2005, *La Traduzione Filmica. Aspetti del Parlato dall'Inglese all'Italiano*, Carocci, Roma.
Peräkylä A., Antaki C., Vehviläinen S. and Leudar I. (eds.) 2008, *Conversation Analysis and Psychotherapy*, Cambridge University Press, Cambridge.
Piazza R. 1999, *Dramatic Discourse Approached from a Conversational Analysis Perspective: Catherine's Hayes's Skirmishes and Other Contemporary Plays*, in "Journal of Pragmatics" 31, pp. 1001-1023.
Piazza R. 2006, *The Representation of Conflict in the Discourse of Italian Melodrama*, in "Journal of Pragmatics" 38 [12], pp. 2087-2104.
Quaglio P. 2009, *Television Dialogue. The Sitcom Friends vs. Natural Conversation*, Benjamins, Amsterdam/New York.
Rossi F. 2003, *Cinema e Letteratura: Due Sistemi di Comunicazione a Confronto, con Qualche Esempio di Trasposizione Testuale*, in "Chaos e kosmos" 4, pp. 1-28. http://www.chaosekosmos.it/pdf/2003_02.pdf (30.03.2018).
Sacks H., Schegloff E. and Jefferson G. 1974, *A Simplest Systematics for the Organization of Turn-taking for Conversation*, in "Language" 50, pp. 696-735.
Schegloff E. 1968, *Sequencing in Conversational Openings*, in "American Anthropologist" 70, pp. 1075-1095.
Schegloff E. 1972, *Notes on a Conversational Practice: Formulating Place*, in Sudnow D.N. (ed.), *Studies in Social Interaction*, Macmillan, New York, pp. 75-119.
Schegloff E. 1980, *Preliminaries to Preliminaries: "Can I Ask You a Question?"*, in "Sociological Inquiry" 50 [2], pp. 104-152.
Schegloff E. 1988, *Goffman and the Analysis of Conversation*, in Drew P. and Wootton A. (eds.), *Ervin Goffman: Exploring the Interaction Order*, Polity Press, Cambridge, pp. 89-135.
Schegloff E. and Sacks H. 1973, *Opening up Closings*, in "Semiotica" 8 [4], pp. 289-327.
Sinclair J. 1991, *Corpus, Concordance, Collocation*, Oxford University Press, Oxford.
Sinclair J. and Coulthard M. 1975, *Towards an Analysis of Discourse. The English Used by Teachers and Pupils*, Oxford University Press, Oxford.
Stokoe E. 2008, *Dispreferred Actions and Other Interactional Breaches as Devices for Occasioning Audience Laughter in Television "Sitcoms"*, in "Social Semiotics" 18 [3], pp. 298-307.
Taylor C. 1999, *Look Who's Talking. An Analysis of Film Dialogue as a Variety of Spoken Discourse*, in Lombardo L., Haarman L. and Morley J. (eds.), *Massed Medias. Linguistic Tools for Interpreting Media Discourse*, LED, Milano, pp. 247-278.

Taibi D., Marenzi I. and Ahmad Q.A.I. in press, *Constructing the House Corpus: Planning Scene Level Indexing and Functionalities*, in Baldry A., Loiacono A. and Bianchi F. (eds.), *Representing and Redefining Specialised Knowledge: Medical Discourse*, ESE – Salento University Publishing.

West C. 1983, *"Ask Me No Questions…" An Analysis of Queries and Replies in Physician-Patient Dialogues*, in Todd A.D. and Fisher S. (eds.), *The Social Organization of Doctor-Patient Communication*, Center for Applied Linguistics, Washington, DC, pp. 75-106.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 425-444
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p425
http://siba-ese.unisalento.it, © 2019 Università del Salento

THE "POWER OF INCANTATION": A LINGUISTIC ANALYSIS OF EUROPEAN THEME PARKS' WEBSITES

SARA CORRIZZATO, VALERIA FRANCESCHI
UNIVERSITÀ DI VERONA

Abstract – Representing "an instance of highly specialized discourse" (Gotti 2006, p. 21), the language of tourism has become an area of great interest in academia, where its multifaceted essence has been studied from different, but intertwined, perspectives. A common feature that characterizes tourist materials is the shared aim of convincing future clients to choose the tourist attraction they are promoting, implementing specific verbal, visual and multimodal techniques. Amongst the strategies commonly associated with the language of tourism promotion tense and magic can be found (Dann 1996). The first projects the tourist outside of present time, denying the existence of time itself, or by transporting the tourist into a nostalgic past or into a utopic future. Magic, on the other hand, relies on the "power of incantation", and on the transportation to locations that do not only exist outside of time, but also outside of a physical place. This contribution aims at investigating how such strategies come into play in the promotion of major theme parks across Europe. Looking specifically at the language used in the construction of narratives in the official websites of eight major European theme parks, the data will be analyzed through both a qualitative and quantitative approach. The use of keywords and expressions pertaining to these two dimensions will be analyzed and commented upon, with the aim of investigating which distinctive promotional strategies are used to "convert [tourists] from potential to actual clients" (Dann 1996, p. 2) in the domain of theme parks.

Keywords: theme parks; language of tourism; keywords analysis; corpus linguistics; discourse analysis.

1. Introduction

Due to its unstoppable progress and widespread influence on inter- and national economies, the travel and tourism industry is nowadays considered as "one of the world's largest industries with a global economic contribution (direct, indirect and induced) of over 7.6 trillion U.S. dollars in 2016".[1]

[1] https://www.statista.com/topics/962/global-tourism/ (12.03.2018).

According to recent research by UNWTO (The World Tourism Organization), tourism appears to be an essential ingredient for our planet's development and prosperity as well as for people's well-being. In terms of socio-economic development, an increasing number of jobs and enterprises have been created thanks to the growing quantity of urban and rural areas that rely on tourism as a source of income. As a consequence, many areas of the world, almost unknown before, have become popular thanks to new tourist flows.[2] Over the last decades, the tourism sector has indeed shown to be one of the strongest and most variable industries, going from 25 million international tourist arrivals in 1950 to 1322 million in 2017.[3] Most notably, in 2016 international tourist arrivals worldwide have risen up to 1235 billion, demonstrating that the top ten international destinations were France, the United States, Spain, China, Italy, the United Kingdom, Germany, Mexico, Thailand and Austria.[4]

Although the tourism sector has been proven by a recent EUROSTAT study[5] to be resilient in face of economic crisis, it "is primarily a services-based industry – the principal products provided by tourism businesses are experiences and hospitality. These are intangible products and more difficult to market than tangible products such as cars" (Davashish 2011, p. 1). For these reasons, promoting a destination is recognized as one of the key elements to attract tourists and to win customer loyalty.

However, promotional strategies connected to tourism have to take into consideration a significant number of factors: "tourism is a volatile business", as Nagle (1999) clarifies, because "not only is it subject to climatic change, it is also influenced by political instability, currency fluctuations, changes on fashion and changes on economy" (Nagle 1999, p. 9). As a matter of fact, social, cultural, historical, political, economic, legal or technological factors, just to name a few, can deeply affect perspective tourists' choices. Yet in 1979, Cohen's words indeed aptly explain the concept that considering tourists "as 'travellers for pleasure' […] is a superficial view of the tourist" (Cohen 1979, p. 179).

[2] UNWTO - Tourism Highlights https://www.e-unwto.org/doi/pdf/10.18111/9789284418145 (10.03.2018).

[3] http://media.unwto.org/press-release/2018-01-15/2017-international-tourism-results-highest-seven-years (12.02.2018).

[4] https://www.e-unwto.org/doi/pdf/10.18111/9789284419029 (14.03.2018).

[5] http://ec.europa.eu/eurostat/statistics-explained/index.php/Tourism_industries_-_employment#Tourism_industries_prove_resilient_to_the_economic_crisis_and_provide_jobs_for_women_and_young_people (13.03.2018).

1.1 Understanding the language of tourism promotion

A closer look at contemporary tourists indeed demonstrates that they are not a monolithic community, but they actually show different characteristics, being driven by diverging motivating factors. In this regard, Dann (1996) recognizes four theoretical perspectives that contribute to delineate tourist profiles: if the so-called authenticity perspective is supposed to illustrate people's desire to live authentic experiences, merging with the real lives of others, the strangerhood hypothesis focuses more on the tourist's "appreciation of the experience of strangeness and novelty" (Cohen 1972, p. 165).[6] The third categorization, defined play perspective, "sees tourism as a play, as a game in which popular pleasure and fun are key concepts" (Cappelli 2006, p. 50). Tracing its roots "in the growth of the British seaside resort […] and its organization of proletarian mass leisure in holiday camps, leaving the middle and upper class to gaze romantically on nature" (Dann 1996, p. 18), this perspective puts its emphasis on spectacle, trying to meet clients' desire for creativity, storytelling, humour, wit, music and fantasy. In Urry's (1990) opinion individuals are attracted by hyperreality, where temporal and special constraints do not exist. Finally, the last categorization, the conflict dimension, has been mostly applied to tourism flows towards less developed countries, as its definition finds its roots in the assumption that interest for this type of tourism grows because of discrepancies existing in wealth distribution and power location.

Offering eye-opening reflections on the role globalization has played in re-shaping the idea of travel and tourism, Fyall and Garrod (2005, p. 3) also point out that traditional promotional methods within business companies are being replaced by "enduring collaborative arrangements", and this recent approach seems to perfectly fit tourism industry, "where fragmented, multi-sectoral and interdependent nature of tourism provides a catalytic focus for inter-organizational co-ordination" (Fyall, Garrod 2005, p. 3). In order to attract more clients, different marketing resources have been created,

[6] The accurate description of tourists within the second category given by Cohen (1979) also helped the following researchers in the field categorize them into five subcategories. Those defined as "recreational" tend to live the experience as a relief, i.e. "a temporary and enjoyable […] movement away from the Centre" (Dann 1996, p. 14). After their playful trip, they will be completely part of their "Centre" again. The second type of tourists are called "divisionary", since they are driven by boredom and dissatisfaction with their "Centre". Unlike the previous example, the so-called "experiential" tourists, dissatisfied with their every-day routine, find relief in unknown places. A more elaborated version of experiential tourism is identified as "experimental", "in which travel is determined by chance and whim, rather than by tourism industry" (Richards 2010, p. 79). The fifth category of tourism is presented as "existential": the individual in fact replaces the "Centre" with the "Other". He completely identifies with the "Other" abandoning his own socio-cultural origins.

improved and exploited throughout the years: Francesconi (2014) offers an accurate classification of "genre families", as defined by Calvi (2010, p. 16), in tourism discourse by recognizing seven categories. The first one, labelled "editorial", includes travel books, travel guides, travel and tourist magazines; the second group, called "institutional", comprises official leaflets, brochures, websites and advertisements. In the third category, defined as "commercial", hotel brochures, leaflets, advertisements and travel agent websites are specified; while tickets, bookings, cards and invoices are examples of the "organizational" group. Within the "legal" dimension, regulations and norms can be found, while critical volumes, essays and articles are part of the "scientific and academic" sector. Travel blogs and travel chats exemplify the last category, called "informal" (Calvi 2010, p. 15).

The heterogeneity of print and digital materials makes their creation really complex, so that several aspects in their construction have to be taken into account. Basically, as Dann (1996, p. 171) suggests, the techniques used are predominately verbal - i.e. the use of figures of speech, the relevant role of key words or ego-targeting – and they generally successfully coexist with visual techniques aiming at optimizing promotional materials' effect through the specific use of colours, format and visual cliché. However, the plurilithic nature of tourism and its heterogeneous interactants has been acknowledged too, so marketing and promotional strategies have become of central interest in order to fully understand how combinations of words and pictures can "persuade, lure, woo and seduce millions of human beings and, in so doing, convert from potential into actual clients" (Dann 1996, p. 2). Marketing experts rely on the aforementioned four perspectives to create promotional materials that could be able to attract future clients by exerting considerable leverage on their desires and needs. According to a general observation, parts of speech become attracting elements aiming at tempting people; therefore, the promotion of a given destination implies a creative and unique approach. Language of authentication, of differentiation, and of appropriation are in fact directly associated respectively to the authenticity, the strangerhood and the conflict perspectives, while linguistic choices defined as "a language of recreation" are those connected to the play perspective (see Dann 1996).

1.2. Theme park promotion

Due to their heterogeneous and multifaceted nature, theme parks are generally associated with various perspectives: if those most common are the play and the strangerhood ones, the authenticity perspective may be also taken into consideration when related to the notion of hyperreality. Indeed, in Dann's words "authenticity does not reside in reality, but in an interpreted representation of reality" (Dann 1996, p. 19). The notion of hyperreality, originally theorized by Baudrillard (1981), has been often associated with places such as theme parks, where "the distinction between reality and fantasy is momentarily blurred" (Brown 2016, p. 24) and the "experiences and settings" guests encounter "are engineered to allow consumers to vicariously experience some other place, time, or reality" (Edvardsson *et al*. 2005, p. 151), to the point that the fictional reality appears to be more credible than the real (Brown 2016, p. 24; Eco 2003, p. 59; Sangiorgi 2012, p. 69). The language of authentication, as will be seen, may be found in theme park promotional material, to highlight its hyperreal nature. This sort of hyperreality is created by rejecting the traditional idea of time and space in favour of a new dimension in which out-of-time and out-of-space magical microcosms can develop; visitors have indeed the possibility to sink in unknown worlds where conventional living rules do not count (Urry 1990). The theme park itself may indeed be seen as a non-place (Augé 1997; see also Sangiorgi 2012, p. 69). In this specific case, promotional language used to develop future visitors' interest aims at "playfully recreate[s] reality" (Dann 1996, p. 21-23). Within a crafted scenery where tourists are invited to play "the role of children by evoking themes of regression and pleasure" (Dann 1996, p. 117), linguistic choices are proven to be of great importance to achieve the goal, which is mainly governed by the "three Fs" model: Fun (Cohen 1985), Fantasy (Boorstin 1987) and Fairy tales (Jafari, Gardner 1991). As a matter of fact, the harmonic co-existence of past, present and future tenses and the persistent reference to magical worlds and fairy tales contributes to shape the visitors' desire to be part of the new environment (see Sangiorgi 2007a, 2007b, 2012). Perceiving the theme park experience as a journey through time and space, timelessness is reinforced so that the present seems to disappear, leaving room to structures pointing to either the past or the future. Past tenses are mainly used to come back to forgotten worlds, whereas future tenses testify a promise to make visitors have fun, laugh and live thrilling and magic experiences. Magic indeed is seen as an element that "misrepresents time in space and vice versa" (Dann 1996, p. 55) captivating and luring clients.

From a more general perspective, the use of language within the field of tourism has become of great interest in the study of promotional strategies and recipients' response, since the lexicon of tourism has been classified as a form of specialized discourse (Gotti 2006). Studies concerning specific linguistic traits and their impact on the tourism sector have indeed focused on a variety of promotional tools, demonstrating that the choice of words can deeply influence the perlocutionary force of utterances not only promoting the target destination but also fostering different processes of socio-cultural representation and negotiation. In some cases, both visual and textual features "re-present the place, […] which is oriented and manipulated by and through promotional aims" (Francesconi 2007, p. 44). Thus, both print and digital promotional materials have been analyzed from different angles (e.g. De Stasio, Palusci 2007; Hallet, Kaplan-Weinger 2010; Agorni 2012; Manca, Bianchi 2013; Teodorescu 2014; Manca 2016; Gotti, Maci, Sala 2017; Malenkina, Ivanov 2018) to investigate how and to what extent language is intrinsically intertwined with marketing strategies.

Websites, whether tourist operator boards, destination websites, or Facebook pages, have recently become the most investigated type of tourism promotion material, as testified by the numerous studies carried out in recent times (e.g. Cesiri 2016, 2017a, 2017b; Bianchi 2017a, 2017b; Pierini 2008, 2009; Manca 2008, 2009, 2013; Maci 2017; Castello 2002, 2017). Linguistic and rhetorical features of the language of tourism online appear to be consistent with those originally identified by Dann (1996) and further developed over the years by other researchers in the field.

2. The present study

The purpose of this paper is to analyze the official websites for major European theme parks and determine which linguistic and rhetorical devices are most frequently used to promote the parks and their services. We expected to find a pervasive use of two of Dann's perspectives, namely the strangerhood and the play perspective, alongside a stronger presence of words related to the dimension of magic. Due to the very nature of theme parks, which may include settings reconstructing entire geographical areas or historical periods, the a-temporal, a-spatial dimensions would also be likely to be stressed in the websites.

More specifically, we would investigate whether these websites promote their destinations by making use of what Dann (1996, p. 55) has called "the power of incantation"; in that respect, he states that the dimensions of time in space are expanded and recreated within the physical boundaries of the park and people are transformed into someone "other than [them]selves" (Dann 1996, p. 55).

2.1. Methodology and data selection

Eight websites of European theme parks were selected for the creation of the corpus. The eight theme parks were chosen according to the annual number of visitors.[7] A deliberate choice was made to exclude from the list theme parks based on pre-existing trademarked products such as *Disneyland Resort*

[7] https://www.statista.com/statistics/194291/attendance-at-theme-and-amusement-parks-in-europe-since-2010/ (04.09.2018).

Paris and *Legoland*, as we assumed these universes would influence the attractions, shows, and accommodation offered as well as language used to promote them. The French park *Puy du Fou* was also excluded due to its focus on historical themes and shows rather than traditional attractions and rides. Data was therefore collected from the websites of the following theme parks:

- *Europa-Park* (EP), Germany
- *Efteling* (EF), The Netherlands
- *Tivoli Gardens* (TG), Denmark
- *PortAventura* (PA), Spain
- *Liseberg* (LB), Sweden
- *Gardaland* (GD), Italy
- *Phantasialand* (PhL), Germany
- *Alton Towers* (AT), United Kingdom[8]

For each website, three sections were selected for analysis, namely

a) rides and attractions

b) shows

c) hotels and accommodations

As the aforementioned categories were deemed those more likely to include detailed descriptions of the park. In addition, these sections were common to all these websites, with the exception of *Liseberg*, which, as a smaller park, does not have internal accommodation. Elements that appeared repeatedly in multiple website pages, like links to other sections of the website, technical or accessibility information on the rides, etc., were purged from the texts. The final corpus totaled at around 45,054 words.

A mixed quantitative and qualitative approach was adopted to investigate the corpus, drawing from the tenets of corpus linguistics for the quantitative part, and discourse analysis for the qualitative section of the analysis. The first part involved the automatic calculation of single-word keywords in the corpus, with the aim of identifying those terms that emerged as significant in theme park promotion. To this purpose, the corpus analysis tool Sketch Engine (Kilgarriff *et al*. 2014) was used. Amongst the numerous reference corpora available on Sketch Engine, *English Wikipedia* was finally selected as, upon reflection, it was considered the most suitable example of

[8] The parks taken into consideration are characterized by different types of attractions aimed at different age brackets, from family rides oriented towards younger children to intense rollercoasters and thrill rides for teenagers and adults. With the exception of *Tivoli Gardens*, which is a traditional luna park, the others are divided into multiple areas, each one with a different theme (e.g. fairytale, Far West, Arab souk, China, etc.) that is reflected in the rides, shows and food joints. All parks except for *Liseberg* also provide accommodation options, some of which evoke the themes found in the park.

online written informative discourse. As we expected specific words pertaining to the individual parks – such as park, area or attraction names – to score high in keyness due to their absence or very low frequency in the reference corpus, a choice was made to set a value of 100 in the "add-N" parameter for keyword extraction, in order to favor the presence of more common words in the keyword list rather than very rare ones (Kilgarriff 2009).

The resulting keywords were then filtered for noise: park and attraction names, as well as terms pertaining exclusively to the area of hotels and accommodation (e.g. *breakfast*, *bunk beds*, *microwave*, etc.) were purged from the final list, alongside words that only occurred in a single website. Out of the 150 keywords identified by the tool, 120 lemmas were eventually included in our final list.[9] The focus was prevalently on lexical keywords – i.e. adjectives, nouns, verbs, adverbs – which constituted a springboard on which to base the qualitative aspect of our research.

Part of speech	Number of Keywords
Nouns	68
Adjectives	34
Verbs	29
Adverbs	2
Pronouns	6

Table 1
Breakdown of keywords by part of speech.

As may be noticed, the sum of the individual elements exceeds the total number of keywords taken into consideration; this is clearly because some of the lemmas in the list occur as different parts of speech (e.g. *ride*, which occurs as both noun and verb) and are therefore counted in two categories

The keywords were then divided according to themes: two macro-themes were identified that were consistent, as expected, with the play and strangerhood perspectives: fairy tales and adventure. Both themes contribute to the distortion of the dimensions of space and time: park guests are plucked out of spatio-temporal reality and become time travelers; in addition, visitors have access to supernatural universes where the laws and constraints of physics are non-existent and a plethora of fantastic, mythological characters come to life.

[9] This final list is reported in Annex 1.

2.2. Fairy tales and Adventures

Many of the identified keywords could potentially fit in both categories, due to the fact that adventure can be considered an element of fairytales, while others can be more easily distinguished as belonging to one or the other genre. These two themes are expressed through multiple aspects in the website, as may be seen from the following keywords:

- Supernatural/mythological creatures: *dragon, giant, ghost, pirate, fairy*
- Adjectives: *enchanting, exciting, magical, spectacular, wild, brave, fantastic, beautiful, adventurous, amazing, incredible, dark*
- Scenic nouns: *tower, jungle, woodland, garden, land*
- Action verbs: *ride, dare, discover, fly, dream, dance, climb*

A search for collocations relating to these keywords suggest that the websites provide accurate and detailed descriptions for the themed areas of the park as well as the themed attractions, as may be seen from the examples below

> (1) The Matterhorn Blitz awaits, and will whisk you off on a *wild* and rugged mountain valley ride. (EP)

> (2) The king calls upon all brave citizens. Slay the *dragon* that terrorises the kingdom! (EF)

Some of these keywords may have received a particularly high keyness score due to the popularity of themed areas and attractions that may be common in multiple parks, such as a fairy tale area, or because it is part of the name of an attraction, or due to the timeframe in which the collection of data took place, that is, around Halloween.

Scenic nouns, which are a common feature of the language of tourism and have shown high keyness scores in previous studies (Kang, Yu 2011, p. 134), are often accompanied in this corpus by adjectives that once again evoke desirable connotations, relating the locations to experiences that would be unimaginable in ordinary life: "beautiful, faraway *lands* with a distinct Asian charm" (PhL), "the ruins of an ancient Aztec *tower* covered in hieroglyphics" (PA), "the treasures of the thick unexplored *jungle*!" (GD) However, the idea of potential danger is also inherent in the narrative presented in the websites, where rides are often presented, as in traditional tales, as obstacles to overcome or enemies to escape from or vanquish. This can be seen through both adjectives and verbs in the corpus, as in example 2 above, or in other instances, such as "*dare* you enter our terrifying scare mazes" (AT), "behold this creepy scene in the *dark* and try not to be afraid" (EF), "blood-curdling encounters with terrifying monsters and creatures of the night" (PA). In the fictional world of theme parks, using this type of negatively-connoted terms does not reflect negatively on visitor interest but, on the contrary, a safe 'scare' or obstacle is welcomed by visitors.

2.3. Time and space

The stress on the dimensions of fairy tales and adventures goes hand in hand with, and contributes actively to, one of the main techniques employed in the websites analyzed, that is, the distortion, to the point of subsuming entirely, of the spatial and temporal dimensions. The disappearance of the conventional notion of time is a common device in promotional tourism discourse (Dann 1996). In the theme park setting, however, this technique appears to be amplified. First, space is also involved, as often multiple geographical settings are reproduced – specific cities – or evoked – countries, or wider areas – within the confines of the park. The guest therefore has the impression of moving from one city – or country – to another in a matter of minutes, as in Europa Park, whose themed areas are inspired by European countries. The extracts below exemplify this technique, as both highlight the opportunity for the guest to travel to the four corners of the planet while being in the park.

> (3) Get ready to experience the greatest adventure for fearless adventurers, whether it means crossing the North Pole or the Jungle, rather than the Wild West or the fascinating Arab world, your exploration will be breathtaking! (GD)
>
> (4) You can travel the world in a few minutes (EF)

Vocabulary pertaining to traveling was present in the keyword list. Words like *ride* and *journey* have high keyness, especially the former as in its noun form it is used to indicate the attractions. Other keywords occurring extensively across parks pertain to the realm of spatial movement, such as *boat, fly, trip, walk, gondola, train, climb, world,* some of which overlap with the dimensions of fairy tales and adventure. The scenic nouns listed in those dimensions also contribute to providing an idea of immediate long-distance travel, as if the guest was being teleported from a location to another. Guests are indeed invited to "hop aboard one of these *boats*" (PA), "witness the *world* from above" (EP) and "float hundreds of feet up in *gondolas* all the colors on the rainbow" (AT)

Passages like "this old steam *train* departs from the Estació del Nord and takes you to the Far West of PortAventura" (PA) connect the erasure of the spatial dimension to that of the temporal dimension. Visitors can move not only through space, but also through time, moving easily through historical eras, from the Far West to Viking Scandinavia or to the Italian Renaissance as guests ride in "one of Leonardo da Vinci's ingenious flying machines!" (EP).

Visitors to theme parks can therefore escape the constraints of time and space during their stay at the park, but, in addition, they can also escape reality in its entirety, as they may find themselves in fictional settings populated by magical creatures and fairy tale characters: park guests “escape from reality” (EF) and can evoke a genie by rubbing Aladdin’s lamp, or take a “virtual journey into the mysterious world of the spirit of the American Natives” (GD).

2.4. Hyperreality and the Strangerhood perspective

The aspects introduced in the previous sections suggest that tourism promotional material for theme parks both underline their hyperreal nature while at the same time making extensive use of the strangerhood perspective (Cohen 1972), with the latter playing a more prominent role even as they overlap in the description of attractions, shows, and themed hotel rooms. Moving constantly across space and time satisfies the guest’s desire of an “experience of strangeness and novelty” (Cohen 1972, p. 165); at the same time, the authenticity perspective is used to make sure that the exotic experience is also realistic. In the websites analyzed, it is often highlighted that the park’s reconstructions are very faithful reproductions of their real world’s counterparts, aiming at providing the guest with a convincing, immersive experience where fiction is undistinguishable from reality. The use of language of authentication, commonly associated with the authenticity perspective (MacCannell 1989), may be interpreted here as stressing the nature of the park as hyperreal, where fictional objects and experiences are perceived as “more real than the real” (Brown 2016, p. 24). This can be seen in the websites, starting from the presence of the adjective *real* in the list of identified keywords, which suggests an equivalence between the fictional world of the park and the actual locations or events it attempts at reproducing. Guests can indeed see “*real* soldiers” (PA), a “*real* Night Garden”[10] (AT), and have a “*real* farm adventure” (EP). The words authentic, typical, original, while not listed among the keywords, are also found, as may be seen in the examples below:

(5) Transport yourself to an authentic hacienda in colonial Mexico (PA)

(6) Observe the typical Nordic wildlife (LB)

(7) Experience the rhythm of Africa for yourself, live! On original drums from West Africa – djembes (EF)

[10] In the Night Garden is a British TV show aimed at children of pre-school age.

The language of authentication is employed here to legitimize the fictional world of the parks, although it has to be noted that this technique is not used consistently in all websites, as it was more common in the Spanish park, with these terms occurring sporadically, or with other meanings, in the other websites. The language of authentication, and more specifically the word *real*, also contributes to the construction of a play perspective – and in turn to the transformation of the guest – which will be investigated in the next section. More common and evident across the corpus is the adoption of a strangerhood perspective, through which the guests leave their world behind and are taken to strange, new, fantastic dimensions.

This perspective is here achieved through a language of differentiation that highlights the element of discovery, mystery and adventure, that is, anything that can offer a respite from the boring routine of home. The spectacular and the impossible are also present, offering unique experiences that would not be available to visitors outside the theme park. We find the language of differentiation applied to existing historical eras and locations – "discover an African village" (EP) as well as to the realms of magic, fairy tales, legends and even science fiction that take the visitor into a completely different dimension. A "mysterious wormhole" (GD), the "legendary city of Atlantis" (EP), and the "humble tower of the Mystery Castle" (PhL) are some of the out-of-the-ordinary settings in which incredible adventures can take place. Once they have "escaped from reality" (EF), visitors will find themselves escaping from "the deadly claws of the tireless predator" (GD), or "a massive volcano eruption" (PA); encountering "curious creatures" (TG), or again going "through lush aquatic landscape" (PhL), and enjoying "spectacular" shows (EF/PA). In the park, childhood dreams can become true: visitors "can live the dream of flying" (EP) and continue living in a dream even after they have woken up in a themed room of a park hotel.

The play perspective is also frequently adopted in theme park promotional material, as the attractions are fictional reconstructions. In such places, tourism is perceived as a game, based on artificial, contrived experiences. The language of differentiation is not only applied to the narrative context through which the rides and shows are presented, but at times the curtain of fiction is parted and the technical characteristics of the rides are exposed in order to attract visitors. Ride information is generally provided for rollercoasters, aimed at visitors looking for a thrill and for whom the magical atmosphere of theme parks goes hand in hand with the adrenaline rush provided by an intense ride. Coasters are described as "spectacular" (EF), "high-speed" (PhL), and having "amazing loops" (TG). Going beyond the identified keywords, comparative forms are used to underline the uniqueness of the ride and the experience it provides: we read of "the highest looping of a catapult coaster in Europe" (EP), "the world's first 14 loop rollercoaster” (AT), the “longest and fastest in Scandinavia” (LB) and so on. The next section will highlight how the play perspective clearly emerges from our corpus through the use of vocabulary, tenses and pronouns.

2.5. The guest as active agent and transformation of the guest

The play perspective, introduced by Urry in 1990, is traditionally associated to theme parks and to the idea of tourism as a game, where the fictional aspect of the holiday experience has a prominent role. This perspective is easily identified in the corpus as emerging from the analysis of multiple linguistic aspects in the websites that pertain to the language of recreation, namely pronoun and lexical keywords, and tense use.

The widespread presence of the personal and possessive pronouns *you*, *your*, which rank respectively first and second for keyness, and the reflexive *yourself* suggest that prospective visitors are addressed directly in the websites. This use of pronouns, known as ego-targeting and also recognized as a common device of the language of advertising (see e.g. Leech 1966), represents the attempt to establish a friendly, informal relationship between the writer and its audience, as the reader perceives that s/he is being addressed individually. Prospective visitors are thus informed of the many opportunities, adventures and experiences they can have at the park, experiences of which they are the active protagonists. Indeed, the visitor is not passively observing what happens in the park or on the rides; on the contrary, the discourse here aims at giving visitors control over their own actions, through which they may influence the course of the plots upon which the rides are based. This strategy fits within the view of the theme park as a hyperreality, where “customers go “on stage” and take an active or passive role in cocreating the experience” (Edvardsson *et al*. 2005, p. 151).

The use of pronouns outlined above contributes to, but is not the only linguistic element employed to, infuse agency in the prospective guest: the use of the verbs *can* and *will*, when accompanying the pronoun *you*, underlines the opportunities for future action that can be undertaken in the park, as may be seen from the following examples:

(8) *you can* put your pirate skills to the test (AT)

(9) *you can* defy waves and sea monsters (EF)

A similar role is played by the imperative form, which is used generously in the websites, as both an exhortation for the prospective guest to take action and as a device to immerse the reader in the narrative and increase desire to try the experience first-hand.

(10) Use the interactive chimes to create a magical tune to woo Rapunzel to safety (EP)

The use of imperatives, however, does not only aim at giving guests the illusion of agency within the park, as in example (10) above, but also have the purpose of transforming them into someone different, someone who exists and lives outside the boundaries of ordinary life.

(11) Hence, get ready to live the life of a curious and... fearless explorer! (GD)

(12) Become a princess in your favourite fairytale (GD)

(13) When the ghosts and ghouls attack, this is your chance to fight back (AT)

Imperatives are not the only verbal form used to create the illusion of transformation into someone different. In the following extract, for example, we see again the use of the future tense through the verb *will*:

(14) You *will* be the Snow Princess in the Kingdom of Ice! (GD)

The use of the future tense in the language of tourism creates in the reader a sense of anticipation and desire. Indeed, here the guest – or, more likely, their children – may be drawn by the promise that they will turn into princesses for the duration of their stay at the parks.

The third element of the language of recreation is constituted by lexical keywords. Namely, the play perspective is achieved through the adjective *real* and the verb to *feel*. As with the elements previously discussed, these also contribute to turning guests into active protagonists, taking different roles according to the chosen ride, show, or hotel. Visitors will become "*real* heroes" (EP), "*real* daredevils" (EF), "*real* princess" (GD). The adjective often appears in combination with *feel*, in the structure "feel like a real", followed by the role taken by the guest: race driver, Red Baron, Cowboy, Far West magnate, and so on. The simile "feel like" also occurs in four websites (AT/EF/EP/PA) to introduce the transformation of the guest; in this case, in addition to roles (e.g. true cowboy, intrepid explorer, real-life adventurers) and fictional characters (e.g. Sinbad the sailor), one ride makes visitors also feel like "a pin ball" (AT). This requires, on the part of the guests, to adopt a suspension of disbelief by which they accept the fictional dimensions created within the borders of the park as reality, shedding their own ordinary identities to take on the roles suggested by the park, embarking on missions, i.e. the rides, and whose actions can determine the success of that mission and the safety of others.

The presence of the play perspective is however not limited to the transformation of the guest; the language of recreation is also used in the description of the experience of the rides, with a preference for thrill rides like roller coasters. As said in the previous section, the narrative is ‘broken’ and an attempt is made to create anticipation in the visitor by evoking the physical sensation of riding roller coasters and other attractions. We find again the keyword *feel* to introduce these passages: passengers on these rides may *feel* dizzy, as though “[they] are being lifted off [their] seat” (PA), or they may feel the pressure of gravity forces and of centrifugal forces. While this aspect may still play into the narrative of adventure that pervades park descriptions, and occur alongside it, it suggests that the websites are also catering to a category of theme park goers that have a specific interest in intense rides and roller coasters. Indeed, both the terms *coaster* and *rollercoaster* are included in the list of identified keywords, indicating the relevance of this type of ride in promoting theme parks online. Hyperreality here coexists with more traditional advertising, where the rollercoaster is perceived and presented as a product, stripped of its magic role in the narrative.

3. Conclusion

As testified by several major studies (see Dann 1996; and Gotti 2006), the language used in the promotion of tourism locations has been collocated within the field of specialized discourses, because it uses specific linguistic strategies aiming at convincing prospective visitors to choose the advertised destination. Verbal – as well as visual – components of promotional materials are essential in the construction of a product which can seduce clients exploiting their (non)-hidden desires. In light of these observations, Dann’s four perspectives accurately clarify different tourists’ needs by highlighting the reasons why they are motivated to select a specific site or travel. The present study demonstrates that the promotion of theme parks does not move away from these shared assumptions, adopting several linguistic strategies that lead even more and more individuals to spend at least one day in a theme park.[11] According to recent research, *Europa Park* (Germany), *Efteling* (Holland) and *Tivoli Gardens* (Denmark) are among the twenty-five most visited theme and amusement parks worldwide in 2016.[12] The keywords identified and selected in the present study show that the parks under examination are presented as a post-modern, hyperreal non-place, where

[11] https://www.statista.com/statistics/194247/worldwide-attendance-at-theme-and-amusement-parks-since-2010/ (05.04.2018).

advertising of the fictional as authentic is combined with the subsuming of the dimensions of time and space and the suspension of the laws of physics to plunge guests into fictional realms in which fun, fairy tales and thrilling adventures are essential elements. The play perspective appears to be the preponderant aspect, with continuous references to the strangerhood perspective. The authenticity perspective may also be included, contributing to the construction of the park as a hyperreal non-place. While the present study focused specifically on the power of incantation and on the magic aspect attributed to theme parks, it should also be noted that a more concrete element is also present in the websites, as testified by the attention to the technical specifications of roller coasters mentioned above, as well as by the numerous filtered keywords pertaining to the services offered by the hotels and other accommodation options within the parks. Nevertheless, official theme park websites appear to make widespread use of the power of incantation and magic, presenting a visit to the parks as a way for all guests – even adults – to live out the fantastic adventures of their childhood dreams.

Bionotes: Sara Corrizzato received a PhD in English Linguistics from the University of Verona (2012). She published *Spike Lee's Bamboozled: A Contrastive Analysis of Compliments and Insults from English into Italian* (Cambridge Scholars Publishing, 2015). Her areas of research include corpus linguistics and LSP, translation theory and audiovisual language, ELT and teacher education. She is a post-doctoral researcher at the University of Verona and is currently teaching English Language and Translation at the University of Verona. She has also taught ESP at the University of Brescia for three years.
Valeria Franceschi is currently a junior researcher in English Language and Translation at the Department of Foreign Languages and Literatures of the University of Verona. She has recently published *Exploring Plurilingualism in Fan Fiction: ELF Users as Creative Writers* (Cambridge Scholars Publishing, 2017). In addition to English as a Lingua Franca, her research areas of interest include digital communication and corpus linguistics.

Authors' addresses: sara.corrizzato@univr.it, valeria.franceschi@univr.it

References

Agorni M. (ed.) 2012, *Prospettive linguistiche e traduttologiche negli studi sul turismo*, FrancoAngeli, Milano.

Augé M. 1997, *L’impossible voyage. Le tourisme et ses images,* Éditions Payot & Rivages, Paris.

Baudrillard J. 1981, *Simulacra and Simulation*, University of Michigan Press, Ann Arbor.

Bianchi F. 2017a, *Strategie promozionali degli operatori del lusso in Facebook*, in “Lingue e Linguaggi” 20, Special Issue, Guido M.G. (ed.) *Strategie di comunicazione dei prodotti di lusso attraverso l'inglese come 'lingua franca' internazionale. Sostenibilità ed emozioni come leve strategiche per lo sviluppo del 'Made in Puglia'*, pp. 239-271.

Bianchi F. 2017b, *The social tricks of advertising. Discourse strategies of English-speaking tour operators on Facebook*, in “Iperstoria” 10, pp. 3-32.

Boorstin D. 1987, *The Image. A guide to pseudo-events in America*, Atheneum, New York.

Brown S. 2016, *Postmodern marketing: dead and buried or alive and kicking?*, in Baker M.J. and Hart S. (eds.), *The Book of Marketing*, Routledge, London, pp. 21-36.

Calvi M.V. 2010, *Los generos discursivos en la lengua del turismo: una propuesta de clasificación*, in “Iberica” 19, pp. 9-32.

Cappelli G. 2006, *Sun, sea, sex and the unspoilt countryside: How the English language makes tourists out of readers*, Pari Publishing, Grosseto.

Castello E. 2002, *Tourist information texts: a corpus-based study of four related genres*, Unipress, Padua.

Castello E. 2017, *Ways of representing and promoting Padua: professional, novice and (non)native voices*, in Gotti, M., Maci S. and Sala M. (eds.), *Ways of seeing, ways of being*. *Representing the Voices of Tourism*, Peter Lang, Bern, pp. 379-400.

Cesiri D. 2016, *Promoting Venice through digital travel guidebooks: a case study of texts written in English and in Italian*, in “CULTUS” 9, pp. 49-67.

Cesiri D. 2017a, *Communicating the Image of Venice: The Use of Discourse Markers in Websites and Digital Travel Guidebooks*, in “Iperstoria” 10, pp. 33-44.

Cesiri D. 2017b, *Balancing Tourism Promotion and Professional Discourse: A Corpus-based Analysis of Digital Travel Guidebooks Promoting Venice in English*, in Vargas Sierra C. (ed.) *Professional and Academic Discourse: an Interdisciplinary Perspective*, “EPiC Series in Language and Linguistics”, EasyChair Publications 2, pp. 247-255.

Cohen E. 1972, *Toward a sociology of international tourist*, in “Social Research” 39, pp. 164-172.

Cohen E. 1979, *A Phenomenology of tourist experiences*, in “Sociology” 13, pp. 179-201.

Cohen E. 1985, *Tourism as a play*, in “Religion” 15, pp. 291-304.

Dann G. 1996, *The language of Tourism. A Sociolinguistic Perspective*, CAB International, Wallingford.

Davashish D. 2011, *Tourism Marketing*, Pearson, Delhi.

De Stasio C. and Palusci O. (eds.) 2007, *The Languages of tourism. Turismo e mediazione*, Edizioni Unicopli, Milano.

Eco U. 2003, *Dalla periferia dell’impero*, Bompiani, Milano.

Edvardsson B., Enquist B. and Johnston R. 2005, *Cocreating Customer Value Through Hyperreality in the Prepurchase Service Experience*, in "Journal of Service Research" 8 [2], pp. 149-161.

Francesconi S. 2007, *English for tourism promotion: Italy in British tourism texts*, Hoepli, Milano.

Francesconi S. 2014, *Reading Tourism Texts: A Multimodal Analysis*, Channel Views Publications, Clevedon.

Fyall A. and Garrod B. 2005, *Tourism Marketing. A Collaborative Approach*, Channel Views Publications, Clevedon.

Gotti M. 2006, *The language of tourism as specialized discourse*, in Palusci O. and Francesconi S. (eds.), *Translating Tourism: linguistic / cultural representations*, Dipartimento di Studi Letterari, Linguistici e Filosofici, Trento, pp. 15-34.

Gotti M., Maci S. and Sala M. (eds.) 2017, *Ways of seeing, ways of being. Representing the Voices of Tourism*, Peter Lang, Bern.

Hallet R.W. and Kaplan-Weinger J. 2010, *Official Tourism Websites: a Discourse Analysis*, Channel View Publications, Bristol.

Jafari J. and Gardner M. 1991, *Tourism and Fiction. Travel and fiction, fiction as a journey*, Centre des Hautes Etudes Touristiques, Université d'Aix-Marseille, Aix-en-Provence.

Kang N. and Yu, Q. 2011, *Corpus-Based Stylistic Analysis of Tourism English*, in "Journal of Language Teaching and Research" 2 [1], pp. 129-136.

Kilgarriff A. 2009, *Simple Maths for Keywords*, in Mahlberg M., González-Díaz V. and Smith C. (eds.), *Proceedings of Corpus Linguistics Conference CL2009*, University of Liverpool, Liverpool. https://www.kilgarriff.co.uk/Publications/2009-K-CLLiverpool-SimpleMaths.doc (12.03.2018)

Kilgarriff A., Baisa V., Bušta J., Jakubíček M., Vojtěch K., Michelfeit J., Suchomel V. 2014, *The Sketch Engine: Ten years on*, in "Lexicography ASIALEX" 1, pp. 7-36.

Leech G. 1966, *English in advertising; a linguistic study of advertising in Great Britain*, Longman, London.

Maci S. 2017, *Meaning-making in Web 2.0 tourism discourse*, in Gotti M., Maci S. and Sala M. (eds.) *Ways of seeing, ways of being. Representing the Voices of Tourism*, Peter Lang, Bern, pp. 131-154.

Malenkina N. and Ivanov S. 2018, *A linguistic analysis of the official tourism websites of the seventeen Spanish Autonomous Communities*, in "Journal of Destination Marketing & Management" 9, pp. 204-233.

Manca E. 2008, *'Immerse yourself in the traditions of the simply way of life': analysing English translations of Italian agriturismi websites*, in "RITT Journal Rivista Internazionale di Tecnica della Traduzione" 10, pp. 33-45.

Manca E. 2009, *Aspetti culturali a confronto: la promozione turistica in Italia e in Gran Bretagna*, in De Rosa G.L. and De Laurentis A. (eds.) *Lingue policentriche a confronto: quando la periferia diventa centro*, Polimetrica, Milano, pp.153-164.

Manca E. 2013, *Describing through the five senses. A contrastive socio-cultural and linguistic analysis of Italian and British tourist websites*, in Manca E. and Bianchi F. (eds.), *Tourism and tourist promotion around the world: a linguistic and socio-cultural perspective*, Salento University Publishing, Lecce, pp. 109-124.

Manca E. 2016, *Persuasion in Tourism Discourse: methodologies and models*, Cambridge Scholars Publishing, Newcastle upon Tyne.

Manca E. and Bianchi F. (eds.) 2013, *Tourism and tourist promotion around the world: a linguistic and socio-cultural perspective*, Salento University Publishing, Lecce.

MacCannell D. 1989, *The Tourist. A New Theory of the Leisure Class*, Schocken Books, New York.

Nagle G. 1999, *Tourism, Leisure and Recreation*, Nelson Thornes, Cheltenham.

Pierini P. 2008, *A Warm Welcome. Guaranteed. Aspetti dell’inglese nei siti degli enti nazionali per il turismo*, in “Studi Linguistici e Filologici Online” 6, pp. 163-202.

Pierini P. 2009, *Adjectives in tourism English on the web: a corpus-based study*, in “CÍRCULO de Lingüística Aplicada a la Comunicación (clac)” 40, pp. 93-116.

Richards G. 2010, *Creative tourism and local development*, in Wurzburger R., Aageson T., Pattakos A. and Pratt S. (eds.), *Creative Tourism. A global conversation*, Sunstone Press, Santa Fe, pp. 78-90.

Sangiorgi S. 2007a, *Postmodern Travels through Standardized Utopias: The Case of Theme Parks*, in Geoffroy, C. and Sibley R. (eds.), *Going Abroad – Travel, Tourism and Migration*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 188-202.

Sangiorgi S. 2007b, *L'ipermercato dei sogni: tecnologia e utopia nei parchi di divertimento a tema*, in “Morus” 4, pp. 193-202.

Sangiorgi S. 2012, *Le parole dell’iperrealtà: la comunicazione dei parchi divertimento a tema*, in Agorni M. (ed.) *Prospettive linguistiche e traduttologiche negli studi sul turismo*, Franco Angeli, Milano, pp. 67-82.

Teodorescu A. 2014, *Tourism Discourse – A linguistic analysis of tourism websites*, in “Language and Literature – European Landmarks of Identity” 15, pp. 367-372.

Urry E. 1990, *The tourist gaze. Leisure and travel in contemporary society*, SAGE Publications, London.

Annex 1. List of selected keywords

Keyword	Keyness		Keyword	Keyness		Keyword	Keyness
you	46.34		stay	5.90		ghost	4.41
your	35.30		fly	5.85		virtual	4.36
ride	31.31		feel	5.82		combination	4.36
fun	20.37		magic	5.71		sit	4.36
experience	20.18		beautiful	5.69		cosy	4.34
train	19.08		adventurous	5.61		everything	4.32
coaster	13.67		thrill	5.50		explorer	4.26
enjoy	13.65		here	5.50		get	4.24
attraction	12.91		comfort	5.50		let	4.24
experience	11.92		tower	5.48		view	4.22
yourself	11.59		garden	5.41		resort	4.19
carousel	11.36		colourful	5.37		splash	4.19
will	11.20		giant	5.33		floor	4.19
our	10.28		interactive	5.23		fast	4.17
holiday	9.61		guest	5.17		walk	4.15
journey	9.47		favourite	5.16		enchant	4.09
roller	9.37		jungle	5.13		pirate	4.09
enchanting	9.04		visitor	5.11		wait	4.09
theme	8.84		water	5.02		family	4.05
playground	8.29		wheel	5.02		real	4.01
exciting	8.29		trip	5.01		entrance	3.99
luxury	8.02		rabbit	4.96		excitement	3.99
boat	7.45		relax	4.96		child	3.98
magical	7.44		bumper	4.93		reality	3.98
spectacular	7.07		whole	4.92		kid	3.95
fairytale	7.04		breathtaking	4.90		height	3.95
unique	6.80		dream	4.89		entertainment	3.93
wild	6.75		perfect	4.88		world	3.91
brave	6.57		woodland	4.84		guarantee	3.90
fantastic	6.55		amazing	4.67		gondola	3.90
dragon	6.49		incredible	4.65		slide	3.87
await	6.48		dark	4.55		exclusive	3.87
spinning	6.47		land	4.54		train	3.85
swing	6.29		atmosphere	4.52		sure	3.84
ready	6.23		everyone	4.48		dance	3.83
dare	6.15		crazy	4.45		climb	3.82
discover	6.09		ghost	4.41		can	3.82
rollercoaster	6.09		virtual	4.36		cars	3.82
little	6.08		combination	4.36		far	3.81
spin	6.07		sit	4.36		sky	3.78
park	6.06		cosy	4.34		indoor	3.78
speed	5.90		crazy	4.45		fairy	3.76

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 445-460
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p445
http://siba-ese.unisalento.it, © 2019 Università del Salento

THE DISCOURSE OF ASSISTED DYING IN THE BRITISH VS ITALIAN NEWS MEDIA[1]

KIM GREGO,[1] ALESSANDRA VICENTINI[2]
[1]UNIVERSITÀ DEGLI STUDI DI MILANO, [2]UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA

Abstract – This paper proposes an investigation into the discourse that surrounds and informs the communication of medically assisted death. The aim is to produce an analysis of assisted dying discourse in the British vs Italian news media to investigate how they construct this type of discourse in terms of actors, ideologies, and cultural aspects. This has been done by collecting a corpus of texts on the topic from English and Italian speaking online quality newspapers. The collection was analysed using quantitative and qualitative methods. The perspective adopted for the qualitative interpretation was that of Critical Discourse Analysis for the socially relevant issues that the topic addresses. Results show how the language used by the media to deal with this bioethical issue reflects that a) in both corpora, actors tend to be the same, including doctors, patients, courts and parliament, especially in the UK, where a law on assisted dying has not been passed yet; b) in both countries, powerful ideologies are at play, principally in the form of Christian views and especially Catholic ones in Italy; c) some cultural differences emerge, especially as regards the diverse legal systems and the specific cases of British or Italian citizens whose stories have contributed to raising awareness on the issue and to influencing the political debate about it in the respective countries.

Keywords: assisted dying; euthanasia; health discourse; discourse analysis; science popularisation.

1. Background

This study expands on previous research carried out on the history of the terminology related to medically-assisted death. Historically, the only term used in Britain to refer to the subject was, until the 1970s-80s, *EUTHANASIA*, in spite of a heated debate on the subject having raged in the country at least since

[1] Both authors contributed to the research behind this study. Specifically, K. Grego is responsible for sections 1 and 4; A. Vicentini for sections 2 and 3. With this article, K. Grego contributes to the national research programme (Programma di Rilevante Interesse Nazionale 2015) "Knowledge dissemination across media in English: Continuity and change in discourse strategies, ideologies, and epistemologies", financed by the Italian Ministry of Education, University and Research for 2017-2019 (nr. 2015TJ8ZAS).

the 1870s (Bok *et al.* 1998, Grego forthcoming). In the same decades, innovative medical research and the ensuing technologies in developed countries resulted in prolonged life expectancy in those societies, sometimes accompanied by good quality of life, sometimes meaning prolonging living with a(n incurable) disease. This explains the emergence of controversial cases and of a societal debate on the termination of life and the philosophical, religious and legal issues surrounding it. As life-supporting technologies and practices developed, so did the lexicon related to them, giving rise to increasingly multi-layered multi-word terms, reflecting the growing complexity of the concepts, e.g. "*WITHHOLDING TREATMENT*" vs "*WITHDRAWING TREATMENT*". The very term *EUTHANASIA*, for centuries a monolithic word referring to anything happening under its wide umbrella, is now understood in multi-faceted ways by bioethics experts, for instance differentiating it according to type or procedure (van Leeuwen, Kimsma 2016, p. 1196):

Types
- Inward–outward (perspective);
- Voluntary–nonvoluntary-involuntary (consent/request);
- Direct–indirect (procedure);
- Active–passive (action).

Procedures
- Withholding treatment (e.g., not installing a ventilator);
- Withdrawing treatment (e.g., stopping a ventilator);
- Sedating (without food or water supply);
- Administering (a lethal drug).

As can be seen from this categorisation (one of many possible ones), the focus can vary from the actor, i.e. the person(s) performing the act, to the action to be carried out, to the subject's request of and consent to it, to the technicalities involved in the action itself.

Not much literature has been published on the specific terminology related to the issue of assisted dying, despite quite a few contributions focusing on the discourse surrounding it. Recent relevant linguistic works with a discursive perspective include Dieltjens *et al*. (2013), van Brussel and Carpentier (2014), Semino *et al*. (2018), while terminological reflections coming from non-linguistic fields include, for example, Jones (2015).

Earlier investigation by the authors reported on the medical and healthcare lexicon as used by specialists, institutions and the media in the UK and US (Vicentini *et al*.2016; Vicentini, Grego 2016; Vicentini, Grego 2017). Further research (Grego, Vicentini forthcoming) confirmed that medically-assisted death tends to be represented in a similar way by English-speaking researchers, institutions and media, who use linguistic expressions connected with assisted dying[2] comparably, except for *EUTHANASIA*, which is mostly used when discussing possible legislation on the issue.

As regards the medical advances and technologies available to support terminal patients, the situation in Italy is similar and consequently so is the public debate on the possibilities of maintaining or discontinuing this type of life-saving treatment/support. The only differences to the UK lie, firstly, in a delayed appearance of a pro-euthanasia movement in Italy, which began in the late 1970s/early 1980s (UAAR 2018) unlike in the UK, where it started in the late 1800s. Once public, though, the debate appeared to be equally fierce in both countries. Secondly, in spite of having focused on the issue more than one century later, in Italy the law on living wills (*BIOTESTAMENTO*, approved January 2018) was approved earlier than in the UK, where living wills have a legal effect according to common law, but are still not regulated by any specific law (WFRDS 2018). Thirdly, the presence and influence in Italy of the Roman Catholic church has historically influenced all public debates on controversial bioethical issues, including life-termination, from a pro-life perspective (e.g. Scalon 2005, Armocida 2015, Nicoli, Picozzi 2017, Vitale 2017). Both the similarities and the differences between these two countries, with their diverse social actors and power relations at play, make it interesting to explore how the debate on assisted dying is reported in the respective national media.

2. Aims, methods and corpus

Building on Grego, Vicentini (forthcoming), this contribution proposes to look at how the discourse of assisted dying is constructed in the British and in the Italian news media in terms of actors, ideologies, and cultural aspects, and to analyse the differences and similarities.

Ideologies seem relevant to analyse, as they are

> representations of aspects of the world which contribute to establishing and maintaining relations of power, domination and exploitation. They may be enacted in ways of interacting (and therefore in genres) and inculcated in ways of being or identities (Fairclough 2003: p.219).

Given the strict relationship existing between language, culture, and ideology, it was also deemed important to shed some light on the similar and different cultural aspects related to the two countries emerging from the corpora, since

> a culture experiences the propinquity of things, how it establishes the *tabula* of their relationships and the order by which they must be considered [...] to reflect

[2] The expressions considered were: "*ASSISTED DEATH*", "*ASSISTED DYING*", "*ASSISTED SUICIDE*", "*END OF LIFE*", *EUTHANASIA*, "*MERCY KILLING*", *RESUSCITATION*, "*DO NOT RESUSCITATE*", *RESUSCITATE*, "*REFUS* TREATMENT*", "*WITHH* TREATMENT*", "*WITHDR* TREATMENT*".

> relations of similarity or equivalence between things, relations that would provide a foundation and a justification for their words, their classifications, their systems of exchange (Foucault 1966/2005, p. xxvi).

The method employed well-established quantitative tools, provided in this case by *SketchEngine* (Kilgarriff, Rychlý 2003). Qualitative tools were used to interpret data from the perspectives of domain-specific discourse (e.g. Gotti, Salager-Meyer 2006; Garzone, Sarangi 2007; Hui *et al*. 2014) and Critical Discourse Analysis (Fairclough 2003, Wood, Kroger 2000; Wodak, Chilton 2005; Eisenhart, Johnstone 2008) for the power relations and social implications that the discourse of assisted dying clearly entails.

The search term *EUTHANASIA*/*EUTANASIA*, considered a particularly relevant keyword to identify assisted dying discourse (Grego, Vicentini forthcoming), was used to retrieve articles from the online archives of some of the main British and Italian quality newspapers: *The Times*, *The Guardian*, *Corriere della Sera* and *La Repubblica*. *The Guardian* and *La Repubblica* are openly centre-left in their stance, while *The Times* and *Corriere della Sera* are traditionally considered to be slightly more on the conservative side.

19 June 2016 – 19 June 2017			
The Times online archive	*The Guardian* online archive	*Corriere della Sera* online archive	*La Repubblica* online archive
7 texts	112 texts	262 texts	72 texts
4,825 tokens	145,637 tokens	171,312 tokens	55,206 tokens
4,107 types	125,458 types	145,085 types	46,277 types
TTR: 0.85	TTR: 0.86	TTR: 0.85	TTR: 0.84

Table 1
The corpus.

The articles retrieved from the four newspapers vary in number, with 112 coming from the *Guardian* but only 7 from the *Times*. The imbalance is possibly due to the *Times* only publishing a digital version of its paper publication, while the *Guardian* continuously issues various online articles a day. Since they nonetheless reflect what was actually published by the four newspapers containing the key term *euthanasia*, the British and Italian corpora seemed sufficiently comparable.

3. Results and analysis

To answer the research question 'how is the discourse of assisted dying constructed in the British and in the Italian news media in terms of actors, ideologies and cultural aspects?', the four subcorpora were subject to a detailed qualitative analysis. The following tables show the linguistic expressions that were employed to represent such actors, ideologies and cultural aspects.

3.1 The British corpus

Table 2 below refers to occurrences from the Times.

UK, *The Times*
*DOCTOR** (f: 27; 5,596 pmw)
*COURT** (f: 17; 3,523 pmw)
*PATIENT** (f: 14; 2,902 pmw)
*LAW** (f: 14; 2,902 pmw)
BELGIUM (f: 9; 1,865 pmw)
POPE (f: 9; 1,865 pmw)
PORTUGAL (f: 8; 1,658 pmw)
NAZI (f: 7; 1,451 pmw)
PARLIAMENT (f: 6; 1,244 pmw)
Individuals linked to specific cases (David, 12yo, Dutch, f: 10; 2,073 pmw).

Table 2
The Times.

With regard to the actors represented in the corpus, the occurrence of *COURT* (and the related term *LAW*) may reflect the UK's absence of legislation on assisted dying, so that individual cases are evaluated as they arise and are reported in the media. This gap is also the reason why people who seek medically-assisted suicide travel to other countries where it is legal (the same seems to emerge from the Italian corpus). The above testifies to the fact that an assisted dying law has not been passed yet; it is still being discussed by/in the medical community. The presence of the term *PARLIAMENT* indeed indicates another protagonist of the debate; it is used when referring to the bill that was presented and rejected three times in the UK (e.g. "A grandfather has called on *parliament* to legislate in favour of assisted dying before ending his own life at Dignitas", Press Association, *Times*, 10 Dec. 2016; "Why *Parliament* rejected assisted suicide", Anon., *Times*, 4 April 2017, emphasis added in both examples). From the cultural viewpoint, the Italian and the British press represent assisted dying using the same terms, such as *COURT/TRIBUNALE*, *LAW/LEGGE* and *PARLIAMENT/PARLAMENTO*, but making reference to divergent law systems and law-making procedures: i.e. in the *Times*, the common law system.

All of the above could also explain why, in the *Times*, *DOCTOR* (the specialist, the expert) ranks high in the frequency list: doctors are the recipient of (legal and/or medical) provisions enabling them to practice euthanasia: “Under a living will, people can *instruct doctors* to withdraw medical treatment” (Gibb, *Times*, 26 Oct. 2016, emphasis added). In the absence of legislation, they must act as the mediators between their patients’ will and the law, for example in preventing acts that could be considered criminal, such as murder or assisted suicide, or in carrying out a person’s living will. As such, their role is fundamental (e.g. “doctors *prevent distress* as the person dies”; “Doctors *have key roles* in palliative care and in suicide prevention” (Anon., *Times*, 4 April 2017, emphasis added)).

Another significant actor is Pope Francis (*POPE*, f: 9), who, as the highest representative of the Roman Catholic church, is one of the strongest opponents of euthanasia. *PORTUGAL* (f: 8) appears in the corpus because a bill to legalise assisted dying was presented in Parliament in 2017, when the “Portuguese church [put its] last hopes in Pope to stop euthanasia” (Lamb, *Times*, 7 May 2017). In fact, Pope Francis was expected to talk about the issue at a mass exactly when a bill on assisted dying (assisted suicide and euthanasia) was being discussed. This was viewed by *The Times* as Catholic interference in political affairs, especially after the country “legalised abortion and same-sex marriage in the past 10 years” (*Ibid*.). The bill was defeated in Parliament on 29 May 2018, although “it secured enough support to ensure continued debate on the issue” (Khalip 2018).

As for any explicit mention of ideologies, the Nazi intent of employing eugenics in order to biologically improve the Aryan race and the Germanic *übermenschen* that was central to their politics are both represented in the corpus: “A panel of experts is to investigate one of Europe’s leading science research institutes to find and purge it of human brain tissue removed from victims murdered during Nazi ‘euthanasia’ and human experimentation programmes” (Hall, *Times*, 7 Jan. 2017).

Finally, a few specific cases also emerge from the data, although they are not as frequent as in the Italian corpus. In particular, there is the case of David, a 12-year-old Dutch boy who wanted to die against his father’s will. The court decided he had the right to commit suicide. The case sheds light on the situation in a country (Holland) where refusing therapies, including artificial nutrition and hydration, has become a right for minors too.

Similar reflections can be applied to the *Guardian* corpus (see Table 3 below).

UK, *The Guardian*
*COURT** (f: 246; 1,689 pmw)
*LAW** (f: 218; 1,497 pmw)
GOVERNMENT (f: 167; 1,147 pmw)
*PARLIAMENT** (f: 101; 694 pmw)
BILL (f: 77; 529 pmw)
*DOCTOR** (f: 70; 481 pmw)
RELIGIOUS (f: 60; 412 pmw); *RELIGION* (f: 18; 124 pmw)
*PATIENT** (f: 42; 288 pmw)
CHURCH (f: 39; 268 pmw)
NETHERLANDS (f: 16; 110 pmw)
BELGIUM (f: 10; 69 pmw)
SWITZERLAND (f: 8; 55 pmw)
Individuals linked to specific cases (George V; f: 8).

Table 3
The Guardian.

Prominent terms from a quantitative viewpoint are *GOVERNMENT* (f: 167) and *PARLIAMENT* (f: 101), which are indicative of the debate that is under way in the UK at the political level (e.g. "The judges recommended that *parliament* should debate the issue before the courts made any decision to change the law"; "it was for national *parliaments* to decide on such a sensitive issue", Bowcott, Sherwood, *Guardian*, 6 Jan. 2017, emphasis added). The high frequency rate of the word *GOVERNMENT* points out that bills on assisted dying, the first step in the process of law making, are quite numerous. If approved by cabinet committees, they are turned into bills that are introduced in a parliamentary session. To become a law, a bill must be passed by both MPs in the House of Commons and peers in the House of Lords: "A private member's bill introduced by Lord Falconer to legalise assisted dying ran out of time in the House of Lords in 2015" (Bowcott, *Guardian*, 22 March 2017).

The *Guardian* also places significant emphasis on the key figure of the *DOCTOR** (f: 70), and on the fact that they have to be allowed by law "to end the life of a terminally ill patient" (Davey, *Guardian*, 16 Sept. 2016). Although *PATIENT* (f: 42) occurs less often than *DOCTOR*, it is qualitatively more significant: patients are the principal actors, as most of the texts are human-interest stories and, as such, patient-centred. The word *ILL* collocates with *TERMINALLY* 55 times, forming a key phrase that indicates that the patient's clinical situation has reached the final stage of a condition. This, however, may not be enough to undergo forms of euthanasia, as other requirements such as a certain span of life expectancy and psychological/mental lucidity must also be met, as the following example indicates:

> In 1997 the US state of Oregon licensed doctors to supply lethal drugs to *terminally-ill patients who met certain conditions* – that they had less than six months to live, had mental capacity and were acting voluntarily (Finlay, *Guardian*, 12 July 2016, emphasis added).

The *CHURCH* (f: 39) is mentioned quite frequently, but the examples carry no particular connotations. Pope Francis is never mentioned in the entire *Guardian* corpus, which can be interpreted as an ideological choice rather than a mere oversight. By not mentioning him, they keep him (and the Roman Catholic church) out of the assisted dying debate.

In addition, focus is placed on three European states where euthanasia has been legalised: the *NETHERLANDS* (f: 16), *BELGIUM* (f: 10) and *SWITZERLAND* (f: 8), which may be taken to reflect the trend of ‘death tourism’, whereby people travel to seek assistance to end their lives. This is analogous to what emerges from the Italian corpus, although Italians opt for Switzerland rather than the other two countries, for reasons of vicinity.

Finally, specific (present and past) cases contribute to pushing forward the debate on assisted dying. Among them is the incident of King George V, who was injected with fatal doses of morphine and cocaine to assure him a painless death.

3.2 The Italian corpus

Table 4 below shows occurrences from *Corriere della Sera*.

Italy, *Corriere della Sera*
LEGGE (f: 140; 2,536 pmw)
*MEDIC** (f: 123; 2,228 pmw)
ASSOCIAZIONE (f: 82; 1485 pmw)
SVIZZERA (f: 82; 1,485 pmw)
*PAZIENT** (f: 68;1,232 pmw)
*PRESIDENT** (f: 60; 1087 pmw)
PARIAMENTO (f: 37; 670 pmw)
BELGIO (f: 34; 616 pmw)
CHIESA (f: 27; 489 pmw)
“*PROPOSTA DI LEGGE*” (f: 21; 380 pmw)
OLANDA (f: 15; 272 pmw)
PAPA (f: 15; 272 pmw)
TESTO (f: 13; 235 pmw)
FRANCESCO (f: 12; 217 pmw)
BIOTESTAMENTO (f: 12; 217 pmw)
Individuals linked to specific cases (Dj Fabo / Fabiano; Luca Coscioni, Piergiorgio Welby, Umberto Veronesi, Marco Cappato, Davide Trentini, Mina Welby, Eluana Englaro).

Table 4
Corriere della Sera.

In *Corriere* (see Table 4 above), reference is made to specialists performing and patients undergoing assisted dying, with *MEDIC** and *PAZIENT** ranking high in the corpus (123 and 68 occurrences, respectively). Doctors still play a pivotal role, taking on specific responsibilities in managing terminal patients, their (palliative) treatment and their living wills (if any). They apparently operate in spite of a gap in the legislation, which is the reason why the press covers a number of specific cases that highlight this situation. The *Corriere* corpus shows that doctors *must* strictly adhere to precise duties:

> Il medico *è tenuto al pieno rispetto delle volontà del paziente* che possono essere disattese, in accordo col fiduciario, qualora sussistano motivate e documentabili possibilità di miglioramento non prevedibili al momento della sottoscrizione (De Bac, *Corriere*, 12 March 2017, emphasis added).

Proof of the lack of a specific law could be that, in the Italian *Corriere*, *SVIZZERA* (f: 82) features quite significantly, as a country where assisted suicide is legal and where most Italians go to look for support in terminating their life.

Religion and the (Roman Catholic) church are quite well represented quantitatively (*CHIESA*, f: 27; *PAPA*, f: 15, *FRANCESCO*, f: 12), and always mentioned in anti-euthanasia contexts. The role of associations emerges from terms such as *PRESIDENT**, which does not only refer here to top positions in politics, but also clearly highlights the relevance of popular movements seeking to exercise bottom-up pressure and to contribute significantly to the political and ethical debate (e.g. "Mina Welby, *presidente* dell'Associazione Luca Coscioni", Rovelli, *Corriere*, 17 Sept. 2016, emphasis added). The term *PARLAMENTO* features in the corpus because a bill on advance decisions (*BIOTESTAMENTO*, f: 12) was being discussed exactly in the period under scrutiny (e.g. "La Camera dei deputati ha approvato il Ddl sul *biotestamento* che va ora al Senato", Redazione online, *Corriere*, 20 April 2017, emphasis added).

In the Italian corpus, specific cases are covered extensively, more than in the UK corpus; famous cases are: Dj Fabo/Fabiano; Luca Coscioni, Piergiorgio Welby, Davide Trentini, Eluana Englaro. This could mean that the discussion on assisted dying tends to reach the masses only occasionally, when striking episodes are covered by the media and are discussed at parliamentary and political levels.

Quite the same observations apply to *La Repubblica* (Table 5 below).

Italy, *La Repubblica*
LEGGE (f: 518; 13,024 pmw)
PAZIENTE (f: 224; 1,308 pmw)
MALATO (f: 214; 1,249 pmw)
SVIZZERA (f: 212; 1,238 pmw)
PRESIDENTE (f: 119; 695 pmw)
ASSOCIAZIONE (f: 101; 590 pmw)
BIOTESTAMENTO (f: 86; 502 pmw)
TESTO (f: 83; 1,249 pmw)
PROPOSTA (f: 74; 432 pmw)
CHIESA (f: 72; 420 pmw) → Vatican (f: 71/72)
GIUDICE (f: 72; 420 pmw) → rulings (f: 39/72)
*EMENDAMENT** (f: 60; 350 pmw)
BELGIO (f: 57; 333 pmw)
PAPA (f: 53; 309 pmw)
FRANCESCO (f: 52; 304 pmw) → Pope (f: 35/52)
RENZI (f: 46; 269 pmw) → bill (f: 5/46)
RADICALE (f: 46; 269 pmw) → political party (f: 37/46)
OLANDA (f: 30; 175 pmw)
Individuals linked to specific cases (Marco Cappato, Dj Fabo / Fabiano, Piergiorgio Welby, Luca Coscioni, Davide Trentini, Umberto Veronesi, Eluana Englaro, Mina Welby).

Table 5
La Repubblica.

The corpus is particularly patient-centred (*PAZIENTE*, f: 224 and *MALATO*, f: 214) and doctors are represented as the ones who have to take decisions concerning their patients in the absence of a law regulating assisted dying, as shown in the following excerpt:

> un medico che si occupava di malati terminali scriveva "...nei loro occhi, negli ultimi giorni della loro vita, ho letto la voglia di farla finita. Qualcuno mi ha chiesto espressamente di aiutarlo in questo senso […]" (Pisapia, *Repubblica*, 7 Oct. 2016).

Consequently, in the *Repubblica* corpus, too, *SVIZZERA* (f: 212) ranks high, and *BELGIO* (f: 57) and the Netherlands (*OLANDA*, f: 30) occur as well, which confirms that Italy also looks at and perhaps even up to these countries for comparison in matters of assisted dying.

A qualitative check confirmed that the church referred to is always the Roman Catholic church. The Pope here is often mentioned by his name only (*FRANCESCO*, f: 52, PAPA FRANCESCO, f: 35). The former Italian Head of Government Matteo Renzi features quite significantly (f: 46) but without any particular connotation. The only political party referred to is *RADICALE* (f: 46): their historical leader, Marco Pannella (d. 2016), fought a hard battle to establish his right to die, which led to a debate about euthanasia in Parliament. The term *PRESIDENT* (f: 119) is used to refer to associations but also to Italy's

Episcopal Conference and Association of Physicians, which means that more and different actors contribute to the debate, compared to both the UK subcorpora and the *Corriere* subcorpus. Specific Italian legal terminology is used for advance decisions (*BIOTESTAMENTO*, f: 86), which again shows that the Italian parliament was in the process of discussing assisted dying: "Si avvicina infatti la votazione alla Camera per il disegno di legge sul *biotestamento* approvato in commissione affari sociali" (Pasolini, *Repubblica*, 18 Jan. 2017, emphasis added). The term *BIOTESTAMENTO*, incidentally, is a 2009 neologism (Treccani 2017, s.v. *BIOTESTAMENTO*) that was first used in *La Repubblica*, summing up the multi-word term "*TESTAMENTO BIOLOGICO*" (literally, "*LIVING WILL*") in a single, obviously more concise word, in line with the Italian tradition of economy in scientific word formation. "*LIVING WILL*", instead, first appeared in the *Indian Law Journal* in 1969 (OED 2016, s.v. "*LIVING WILL*") as a denomination strategy: "The document indicating […] consent may be referred to as 'a *living will*'". This highlights how the term was initially attested in a specialised US publication in the late 1960s, while in Italy it appeared in a popular source only a few years ago. The terms *BIOTESTAMENTO* and *EMENDAMENTO* are naturally employed with reference to the Italian political and legal system, which is based on civil law.

4. Discussion and conclusions

In terms of actors and ideologies, in the UK corpus, the actors and protagonists are divided into two groups, both influencing and developing the debate: doctors/patients on one side and campaigners/legislators on the other. These still appear to be at the centre of the discussion because, in spite of the British public debate on euthanasia having a long history, with campaigns in its favour starting from the 19th century and the first bill reaching Parliament in 1936 (only to be rejected), the country still lacks a specific law on the subject. The high frequency of the term *COURT* (f: 17 *Times*, 246 *Guardian*) testifies to the fact that a legislation on assisted dying is still missing, the court being the place where specific and problematic cases are discussed individually.

Considering ideologies, the strongest one emerging from the British press seems to be the one informing the Christian (both Catholic and Anglican) position on euthanasia, yet the ideas of the Christian representatives reported in the press are not necessarily against it: "The Catholic right sees this as its last bastion" (Lamb, *Times*, 7 May 2017), "Desmond Tutu: I want right to end my life through assisted dying" (Sherwood, *Guardian*, 7 Oct. 2016). Other ideologies hinted at are the Nazi eugenics programme, which again shows how the issue has been and still is controversial. The Nazis' horrific deeds are, indeed, one of the arguments frequently brought forward by anti-euthanasia supporters.

As for the Italian corpus, the debate is constructed by and around very much the same actors. However, in Italy the public debate on the subject is a more recent one, starting in the late 1970s/early 1980s: no bills on active euthanasia proper have been passed or rejected as yet, although a law on advance decisions was approved on 31 January 2018. The Italian corpus seems to reflect the heated debate that was ongoing at the time of collection, with bottom-up pressure from pro-euthanasia associations, thus explaining the prominence of terms such as ‘battaglia’, ‘lotta’, etc. (for a review of end-of-life metaphors, see Semino *et al*. 2018). As emerges from the analysis, the radical party is the only supporter of active euthanasia at the political level while, at the other end of the dialogical spectrum, stand the Pope and the Roman Catholic church. Although a first law on advance decisions now exists in Italy, the controversy is still ongoing, and the same actors and ideologies appear to be at play.

Furthermore, two different approaches to assisted dying issues and legislations seem to emerge from the analysis, possibly deriving from and pointing to two diverse cultures and legal systems. The UK media, to begin with, mainly look at Belgium and the Netherlands as countries where forms of euthanasia are legal, possibly for their geographical proximity. The Pope and the church, not only the Roman Catholic one but also the Protestant one, appear to be referred to and seen as opponents of any move towards an assisted dying law.

The Italian media, on the other hand, seem to represent the Italian people as looking at and even up to Switzerland for its laws on assisted dying, and as the place to go, closer than Belgium and the Netherlands, to get support in ending one’s life medically. The focus of the debate in the period analysed is on the bill (“*TESTO DI LEGGE*” or just *TESTO*) being discussed then, which was subsequently passed and turned into a law at the beginning of 2018. The Pope always stands out as a reference point in the discussion, most likely for the central role the Roman Catholic church has in Italy’s culture, which is also due to the geographical position of the Vatican within the city of Rome.

In the presentation of arguments in favour and against assisted dying, the two Italian newspapers make ample recourse to human-interest stories, or individual cases of specific people (e.g. Luana Englaro, DJ Fabo, etc.), often appealing to the emotions of readers. In other words, it seems that, in order to make the Italian public reflect upon these issues, the human-interest factor needs to be presented first. This could be due to the fact that, possibly owing to the strong influence of Catholic beliefs, the debate on the matter is relatively recent, only beginning in the late 1970s/early 1980s, along with the ones on divorce and abortion, thus determining an emotional rather than rational approach to the problem.

Although this is a limited qualitative analysis, it seems to show that this is still a multi-layered and very controversial topic, with premises, perspectives, and implications that must take into account cultural factors on top of strongly rooted ideologies informing powerful social actors (an interesting result that emerged was the use of *PRESIDENT* to refer to both institutions and citizens' associations). Nonetheless, precisely for this reason, the debate does not look as though it will provide widely shared answers in terms of legislation, at least not in the near future. That is also why it appears very interesting to keep observing and researching the discourse of assisted dying procedures and how they are regulated by different countries worldwide, bearing in mind that this discourse may be taking different paths to provide different solutions as ethics and ideologies change and adjust to a wider, global view of problems that nevertheless affect all humans indistinctly.

Bionotes: Kim Grego is Assistant Professor of English Language and Translation at the University of Milan, where she teaches English Language and Linguistics. Her interests include Translation Studies, ESP, Critical Genre Analysis and Critical Discourse Studies. She leads the University of Milan's research unit of the "Age.Vol.A. - Ageing, Volunteers, Assistants - Multilingual tools for assisting the ageing" project, funded by Fondazione Cariplo (2018-2021). Her recent publications include: "Ethical Aspects in Web Marketing for Seniors" (2018) and "Age-specific Tourism: Representations of Seniors in the Institutional Discourse of Tourism" (2017).
Alessandra Vicentini is Associate Professor of English Linguistics and Translation at the University of Insubria, where she teaches English Language and Linguistics. Her research interests include ESP (Medicine/Healthcare, Bioethics, Media), Discourse and Genre Analysis, English and Anglo-Italian historical grammaticography and lexicography. She is the principal investigator of the "Age.Vol.A. - Ageing, Volunteers, Assistants - Multilingual tools for assisting the ageing" project, funded by Fondazione Cariplo (2018-2021). Recent publications are: *Stem cells and (pseudo)science: Discursive aspects of the Stamina case as seen in Nature*, 2018; *'Meat gives you cancer'. The popularisation of scientific news with public health relevance*, 2018.

Authors' addresses: kim.grego@unimi.it, alessandra.vicentini@uninsubria.it

References

Armocida G.O. 2015, *Eutanasia: una questione diversamente problematica nella storia della medicina*, Le Monnier, Firenze.
Bok S., Dworkin G. and Frey R. 1998, *Euthanasia and Physician-Assisted Suicide: For and Against*, Cambridge University Press, Cambridge.
Dieltjens S., Heynderickx P., Dees M. and Vissers K. 2013, *Linguistic Analysis of Face-to-Face Interviews with Patients with An Explicit Request for Euthanasia, their Closest Relatives, and their Attending Physicians: the Use of Modal Verbs in Dutch*, in "Pain Practice" 14 [4], pp. 324-331.
Eisenhart C. and Johnstone B. 2008, *Discourse Analysis and Rhetorical Studies*, in Eisenhart C. and Johnstone B. (eds.), *Rhetoric in Detail: Discourse Analyses of Rhetorical Talk and Text*, Benjamins, Amsterdam/Philadelphia, pp. 3-21.
Fairclough N. 2003, *Analysing Discourse: Textual Analysis for Social Research*, Routledge, London/New York.
Foucault M. 1966/2005, *The Order of Things*, Routledge, Abingdon.
Garzone G. and Sarangi S. (eds.) 2007, *Discourse ideology and specialised communication*, Peter Lang, Bern.
Gotti M. and Salager-Meyer F. (eds.) 2006, *Advances in Medical Discourse Analysis: Oral and Written Contexts*, Peter Lang, Bern.
Grego K. and Vicentini A. forthcoming, *Terminal terminology. The lexicon of medically assisted death across domains of specialisation*. Presented at 21st Conference on Language for Specific Purposes 2017, Interdisciplinary knowledge-making: challenges for LSP research, Norwegian School of Economics, 29 June-1 July 2017.
Grego K. forthcoming, *From euthanasia to end-of-life issues. Changing names for changing concepts*, First international conference on historical medical discourse CHIMED-1, Università degli Studi di Milano, Milan, 14-16 June 2017.
Hui D., Nooruddin Z., Didwaniya N., Dev R., De La Cruz M., Kim S.H., Kwon J.H., Hutchins R., Liem C. and Bruera E. 2014. *Concepts and definitions for "actively dying," "end of life," "terminally ill," "terminal care," and "transition of care": a systematic review*, in "Journal of Pain and Symptom Management" 47 [1], pp. 77-89.
Jones D.A. 2015, *Assisted dying: law and practice around the world*, in "BMJ" 351:h4481.
Khalip A. 2018, *Portugal parliament rejects legal euthanasia in divisive vote*, 29 May 2018. https://www.reuters.com/article/us-portugal-euthanasia/portugal-parliament-rejects-euthanasia-legalization-idUSKCN1IU2DJ (20.08.2018).
Kilgarriff A. and Rychlý P. 2003, *SketchEngine*, Lexical Computing, Brighton. https://www.sketchengine.eu/.
Nicoli F. and Picozzi M. 2017, *Killing or allowing someone to die: a difference defined exclusively by the criteria of "terminal"? Making decisions regarding a patient's death*, in "Medicina Historica" 1 [1], pp. 41-47.
Scalon R. 2005, *Bioetica e sfera pubblica. Un'analisi sociologica del dibattito su fecondazione assistita e eutanasia*, Franco Angeli, Milano.
Semino E., Demjén Z., Hardie A., Payne S., Rayson P. 2018 (eds.), *Metaphor, Cancer and the End of Life. A Corpus-Based Study*, Routledge, Abingdon.
UAAR - Unione degli Atei e degli Agnostici Razionalisti 2018, *Eutanasia*, https://www.uaar.it/laicita/eutanasia/ (20.08.2018).
van Brussel L. and Carpentier N. (eds.) 2014, *The Social Construction of Death. Interdisciplinary Perspectives*, Palgrave, London.

van Leeuwen E. and Kimsma G. 2016, *Euthanasia: International Debate*, in ten Have H. (ed.), *Encyclopedia of Global Bioethics*, Springer International Publishing, Cham, Switzerland, pp. 1194-1202.
Vicentini A. and Grego K. 2017, *Medical web encyclopaedias: linguistic aspects and ethical implications*, "EXPRESSIO" 1 (2017), pp. 149-169.
Vicentini A. and Grego K. 2016, *Vaccines don't make your baby autistic: arguing in favour of vaccines in institutional healthcare communication*, in Mohammed D. and Lewinski M. (eds.), *Argumentation and reasoned action*, Vol. 2, College Publications, London, pp. 999-1020.
Vicentini A., Grego K. and Canziani T. 2016, *A Matter of Terminology, When Terminology Matters: Naming Common Genetic Diseases*, in Garzone G., Heaney D. and Riboni G. (eds.), *Language for Specific Purposes: Research and Translation across Cultures and Media*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 200-218.
Vitale A.R. 2017, *L'eutanasia come problema biogiuridico*, Franco Angeli, Milano.
WFRDS - The World Federation of Right to Die Societies 2018, "Living Wills in the UK", https://www.worldrtd.net/living-wills-uk.
Wodak R. and Chilton P. (eds.) 2005, *A New Agenda in (Critical) Discourse Analysis*, Benjamins, Amsterdam/Philadelphia.
Wood L. and Kroger R. 2000, *Doing Discourse Analysis*, Sage, London.

Cited articles from the corpora

Anon. 2017, *Why parliament rejected assisted suicide*, in "The Times", 4 April 2017.
Bowcott O. 2017, *Terminally ill former lecturer challenges UK ban on assisted dying*, in "The Guardian", 22 March 2017.
Bowcott O. and Sherwood H. 2017, *Terminally ill UK man launches legal challenge for right to die*, in "The Guardian", 6 January 2017.
Davey M. 2016, *Voluntary euthanasia support grows in Victoria, but Turnbull won't back it*, in "The Guardian", 16 September 2016.
De Bac M., 2017, *Biotestamento è legge, sì del Senato. Cosa prevede, quali i punti critici*, in "Corriere della Sera", 12 March 2017.
Finlay L.I. 2016, "Assisted dying: what can the UK learn from places where it is legal?", in "The Guardian", 12 July 2016.
Gibb F. 2016, *End-of-life cases must go to court*, in "The Times", 26 October 2016.
Hall A. 2017, *Archive holds human brain tissue from Nazi era*, in "The Times", 7 January 2017.
Lamb C. 2017, *Portuguese church puts last hopes in Pope to stop euthanasia*, in "The Times", 7 May 2017.
Pasolini C. 2017, *Biotestamento: discussione più vicina. Tagliati 3000 emendamenti*, in "La Repubblica", 18 January 2017.
Pisapia G. 2016, *Il piccolo Davide e il diritto a una morte dignitosa*, in "La Repubblica", 7 October 2016.
Press Association 2016, *Man urges MPs to back assisted dying before ending his life at Dignitas*, in "The Times", 10 December 2016.
Rovelli M. 2017, *Eutanasia, in Belgio il primo caso al mondo su un minore*, in "Corriere della Sera", 17 September 2016.

Sherwood H. 2016, *Desmond Tutu: I want right to end my life through assisted dying*, in "The Guardian", 7 October 2016.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 461-480
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p461
http://siba-ese.unisalento.it, © 2019 Università del Salento

(TRANS)GENDER IN THE NEWS: SPECIALIZED LANGUAGE IN THE UK PRESS
A corpus-based discourse analysis

ANGELA ZOTTOLA
UNIVERSITY OF NOTTINGHAM

Abstract – In recent years, gender variant, non-binary and queer identities have become prominent topics of discussion in society as much as in the field of cultural and linguistic studies, and language has been playing a seminal role in the process of shaping and negotiating these identities. The press, in particular, works as one of the most active agents in the creation of discourses linked to those communities. This paper presents the final outcomes of a research project that focuses on the representation of transgender people in the British press as social actors (van Leeuwen 1996, 2005, 2008) and formulates hypotheses with regard to the use of LSP (Language for Specific Purposes) in this type of representation.

Keywords: transgender identities; CBDA; LSP; British press.

It's not the news that makes the newspaper,
but the newspaper that makes the news.
(U. Eco "Numero Zero", 2015,
Trad. en., p.70)

1. Introduction

The aim of this article is to critically address the role of the British press in the representation of transgender people, considering the divide between popular and quality press. The rationale for this investigation lies in the degree of involvement of newspapers in the understanding and dissemination of issues related to trans identities. In fact, as Fairclough (1995, p. 2) argues, the media influence "knowledge, beliefs, values, social relations [and] social identities".

Scholars of News Discourse and language tend to agree on the fact that "[…] news is socially constructed. What events are reported is not a reflection of the intrinsic importance of those events, but reveals the operation of a complex and artificial set of criteria for selection" (Fowler 1991, p. 2). The construction of news is realized though language, which is never neutral or a casual choice but always a highly constructed and elaborated conveyer of significance.

Due to the sensitive nature of the topic under investigation and to the central role the use of language has in the understanding of transgender identities, it is necessary to adopt a standard for inclusive and non-discriminatory language, as will be demonstrated in the following sections. In fact, when it comes to transgender identities, considerations about word choice or proper grammar use influence and modify the understanding of the identities. It is of the utmost importance to comply with an established and acknowledged use of language, approved and shared by the in-group community. Various guidelines offered by organizations that focus their work on improving transgender people's lives are taken into consideration, among these: GLAAD (https://www.glaad.org); the National Center for Transgender Equality (http://www.transequality.org); All About Trans (http://www.allabouttrans.org.uk); and the Beaumont Society (https://www.beaumontsociety.org.uk/). The perspectives on the use of terminology referring to transgender people are many and vary from group to group; the terminology used here follows the GLAAD guide. These GLAAD guidelines were chosen because they are the most exhaustive and GLAAD specialize in monitoring language use in media. The broad spectrum in which transgender identities are set, together with the understanding that the performative nature of gender identities (Butler 1990), also relates to the concept of transgender community as a social construct that cannot be representative of all transgender people for the very fluid nature of this identity.

The work presented in the following sections looks at transgender individuals as social actors – that is to say, at how the press represents transgender people as "*participants* of social practices" (van Leeuwen 2008, p. 23). The aim of this approach is to shed light on the linguistic practices to which audiences are exposed daily and that consciously and unconsciously work in the background of their minds, shaping their knowledge of issues related to transgender identities. The study also aims at establishing the extent to which terminology related to transgender identities can be considered a type of Language for Specific Purposes.

Sections 2 and 3 will introduce the theoretical and methodological frameworks employed in this work. Section 4 will discuss the analysis and the results emerging in this study, while Section 5 will attempt to draw some conclusions.

2. The language of the press and the representation of (trans)gender identities

2.1. News Discourse in the UK: The quality and the popular press

One may consider the language of the press as a discourse of its own, with its peculiar features and functions, therefore a type of language for specific purposes. The language used in the press reveals a lot about the news itself and the political and ideological standpoint of the newspaper supporting it. Certain editorial choices by newspapers affect attitudes and affairs in society and create stereotypes and biases in culture (Bell 1991, pp. 3-5). Newspaper articles are the main objects of this investigation; thus, it is necessary to describe some of the specificities of this genre.

The main function of the press is to report news. Bednarek and Caple (2012) suggest that one might approach the study of news discourse from a variety of perspectives, including not only linguistics, but also journalism, communication and media studies. Some of the first studies on news discourse focus on the structure of news' language by looking at genre, register and style. This is the case of, e.g., Crystal and Davy (1969), Carter (1988) and Ghadessy (1988), who focus on lexical choices in news articles, with the aim of defining the language of newspaper discourse. Through a sociolinguistic perspective, Bell (1991) and Jucker (1992) address the link between linguistic structures and social context in news articles. The same approach, but with a diachronic perspective, is chosen by Conboy (2010). Other scholars focus on news values by combining the study of evaluation with corpus linguistics (e.g. Bednarek 2006). The latter approach has increasingly become popular in the analysis of news discourse, and will be adopted in this study. The investigation of News Discourse, through the use of different theoretical and methodological frameworks, has also proved to be an effective instrument to uncover patterns in the representation of minorities, particularly of different social groups and ethnicities (see for e.g. van Dijk 1988 on racism in the press; Baker *et al.* 2013 on the representation of Islam; and Partington 2015 on the representation of the Arab world).

In Great Britain, the first daily newspaper appeared in 1702, at the time known as *The Daily Courant* (Allan 1999, pp. 8-11). Since then, the press has seen many changes, but one feature remains consistent throughout the centuries: the British press has maintained its own internal categorizations, from the 'broadsheet' vs. 'tabloid' classification, to the 'quality' vs. 'popular' dichotomy, moving to the 'up-', 'mid-' and 'down-market' newspaper distinction. Newspapers in the UK are also classified according to their political sympathy (left and right leaning), based on the frequency of appearance (daily, weekly, Sunday editions) or the geographical area they cover (nationals and locals) (Seymour-Ure 1996 p. 27).

In 1983, Harry Henry proposed a type of classification that divides periodicals in up, mid- and down-market newspapers. This categorization is specifically based on the socio-economic status of the readership. However, today, more newspapers are read across social classes than ever before (Jucker 1992, pp. 48-58). This adjustment is due not only to the increase in the level of education of all social classes, but also to the growing trend in the use of on-line platforms. In light of this, the up-, mid-, and down-market classification results is no longer accurate and is not taken into consideration in this study.

Similarly, the ‘broadsheet’ vs. ‘tabloid’ dichotomy, inspired by the size and format of the various newspapers (Jucker 1992, p. 48) struggles to survive following the standardization in size of most British newspapers.

A further categorisation system, relevant to this study, is the dichotomy between ‘popular’ and ‘quality’ newspapers. “The quality-popular distinction was sharply drawn by 1945” (Seymour-Ure 1996 p. 27), although the differences in newspapers were already in sight a decade before, since the success of the *Daily Mirror* in 1935, which soon became the first massively distributed popular paper. This distinction is mainly based on the different ways that newspapers chose to prioritize information, and on the linguistic means used to report them.

The differentiation between ‘popular’ and ‘quality’ press seems to have always existed. Conboy (2002, p. 5) claims that the term ‘popular’ in relation to the press closely relates to the definition of popular culture. Here, he refers to Raymond Williams (1976), who points to the origin of the word used to indicate something “[...] ‘belonging to the people’ but also [*carrying*] implications of ‘base or low’”. Conboy (2007, p. 7) posits that the popular press can be defined as a “[…] a set of discourses which establish elements of authenticity in part through its rhetoric and is thus able to establish an inclusivity based on its appeal to wide sections of ordinary people”. The popular press owes its large number of followers to the fact that it was able to attract a “largely working-class readership because of its commitment to delivering a form of journalism these readers wanted to see at a price that they could afford” (Allan 1999, p. 13). According to Fowler (1991, p. 39), generally speaking, the popular press uses “colloquialisms, incomplete sentences, questions and a varied typography suggestin*g* variations of emphasis, [where] the written text mimics a speaking voice, as of a person talking informally but with passionate indignation”. Moreover, its style includes puns, easy to understand terms and basic sentence structure, while the quality press has a more formal type of language.

Among the characteristics proper of the genre, it is possible to highlight that popular newspapers usually have shorter articles, focus more on national stories, and prefer to cover news stories that deal with gossip and the lives of celebrities (Baker *et. Al.* 2013; Conboy 2006). Quality newspapers, by contrast, have a broader area of coverage. Articles in this type of newspapers generally include stories and news related to foreign politics, business and economics, as well as updates from all over the world; the style of writing incorporates specialized terminology and complex sentence structure (Baker *et. al.* 2013; Jucker 1992).

While the aim of all newspapers is more or less similar, i.e. to inform and entertain their audience, specific categories of newspapers assign different priorities to the two aims (Jucker 1992, p. 2). Those newspapers falling under the quality press categorization, such as *The Guardian* or *The Daily Telegraph*, focus on the informative function; while newspaper categorized as belonging to the popular press, like the *Daily Mail* or *The Sun*, give more importance to the entertainment function.

Nowadays, one of the most reliable bodies in charge of keeping track of all newspaper publications, rates of distribution and development of newspaper outlets is the Audit Bureau of Circulation.[1] Information offered by this body has also been taken into consideration in the creation of the dataset under investigation in this study.

2.2. (Trans)Gender identities in the press

As Litosseliti (2002, p. 136) argues, "[n]ewspapers are a prime public site for moral arguments and for constructing values and ideologies". For this reason, news discourse is a great source for the analysis of how ideologies, power relations, and the cultural values of a society are expressed and represented through language. An analysis of the language of the media is useful to identify those discourses that pervade, influence and shape the way people see and understand society, its beliefs and values.

One topic that has gained more and more importance, in this sense, is the study of how gender identities are represented in news discourse. The first issues about gender and language in relation to news discourse analysed by linguists are concerned with the representation of women in the press, and with the percentage of women actually working with and using this means of communication. Carter, Branson and Allan (1998) wrote and published *News, Gender and Power*, a volume which retraces the role of women in the press from different perspectives, from the actual contribution by women in the press to how they are represented and why. Byerly and Ross (2006) published *Women and Media: A Critical Introduction*, extending the study from journalism to media as a more general field of inquiry. Overall, while the linguistic representation of males and females in news stories has been

[1] http://www.abc.org.uk/

investigated for many years (see also Clarke 1992; and Adampa 1999), that of sexual and gender identities has been considered a topic of scholarly analysis only more recently.

At the beginning of the 21st century, several linguists show their interest in how language might represent non-heteronormative/homonormative gender identities. In 2005, Baker looks at discourse prosodies in two British tabloids in relation to the representation of homosexuality and gay men. He concludes that while both *The Mirror* and *The Daily Mail* newspapers relate gay identity to more negative discourses such as that of a "gay lobby' trying to control the world, *The Mirror* does not frequently discuss politically based discourses in relation to homosexuality, while the *The Daily Mail* focuses more on equality and on how gay propaganda influences children. Similarly, Gouveia (2005) analyses the representation of gays and lesbians in a Portuguese newspaper, from a set of articles published across one week with the heading "gay power" in 2001. In his work, Gouveia argues that homosexuality is a topic rarely mentioned in the press and still considered a taboo. The articles he analyses mirror the discussion about civil unions, on which the Portuguese parliament passed a law just few weeks before. According to the author, despite the clear intention to refrain from discriminating against homosexual people, "there [was] also the construction of a sense of fear associated with homosexuality, via the assertion that gays and lesbians have more social and political power than one would expect" (Gouveia 2005, p. 247). This argument brings forward again the idea of the "gay lobby" observed in Baker's study (2005). Morrish and Sauntson (2007) look at the representation of Peter Mandelson, a gay British politician, in British broadsheets, concluding that not only is he portrayed as effeminate but also in a negative way through derogatory stereotypes generally associated with gay men. Hidalgo-Tenorio and Bartley (2015; 2016) conduct two different studies on a corpus of newspaper articles collected from the Irish press following the issuing of a law that recognizes same-sex marriage. The 2015 study focuses on the verbal processes associated to the terms *gay**, *homo**, *lesbian** and *queer**, that is to say, to transitivity processes and how they were used to represent queer identities. The study concludes that the most frequent verbal process used in this case were 'material'. To put it simply, in their data, queer people are mostly represented as active participants. The corpus also shows strong homophobic discourse, where gay men are demonised, victimised and represented as curable patients, while lesbians seem to be generally more accepted by society. In the 2016 study, the aim is to examine how the perception of homosexuality was filtered through the press (Hidalgo-Tenorio, Bartley 2016, p. 9). Among other results, the study highlights that gay people are mostly represented as a community, a group of people working together.

With the exception of the 2015 work by Hidalgo-Tenorio and Bartley, which includes the word 'queer',[2] the studies presented so far focus on homosexual people, and thus discuss LGBT+ identities only to a certain extent. The representation of transgender identities in the news has been investigated by linguists only more recently, in the past few years. An influential work for this research is a study by Paul Baker, published in 2014. In his work, Baker collects a corpus of newspaper articles from the year 2012 and, through the use of corpus linguistics tools, analyses the way the British press depicted trans people in that year. One of his major findings is that:

> The analysis did find a great deal of evidence to support the view that trans people are regularly represented in reasonably large sections of the press as receiving special treatment lest they be offended, as victims or villains, as involved in transient relationships or sex scandals, as the objects of jokes about their appearance or sexual organs and as attention seeking freakish objects. (Baker 2014, p. 233)

Baker also added that some positive representations were retrieved but were less frequent. Zottola (2018a, 2018b, forthcoming), expanding on Baker's work, analyses the semantic prosody carried by the linguistic choices adopted by newspapers to represent transgender identities in the British press. Zottola (2018a, 2018b) discusses the representation of transgender people in the press in relation to lexical choices that do not always support an inclusive use of language for transgender people. Zottola (forthcoming) tackles different semantic categories of representation used to talk about transgender identities, such as, among others, the world of celebrities or collective representation using concepts such as 'community'. Ferraresi (2018) also writes about the press and its lexical choices related to the representation of transgender people. He focuses on citizen journalism, investigating the extent to which negative stereotypes about transgender people are reversed. Finally, Gupta (2018) contributes to the literature on the representation of transgender identities in the press through an analysis of misgendering practices through pronoun use in news reports on Lucy Meadows, a British teacher who was found dead in her house short after returning to her job in her identifying gender.

[2] "Queer" is a more general term referring to all those identities which do not conform to the heteronormative and binary definition of gender.

3. Language for Specific Purposes (LSP) and Corpus-Based Discourse Analysis

3.1. LSP and identity representation

Although Language for Specific Purposes (LSP) is generally regarded as a type of language used by experts communicating within their area of expertise (Bergenholtz, Tarp 1995), it can also be defined as an array of conventions and rules – sometimes overlapping with those used in general language – that are characterized by specific peculiarities such as topic, communicative goal, audience of reference or context of appearance (Cabré 2003). Bearing in mind the discussion about the language of newspapers introduced in Section 2, it is possible to retrieve a number of these features in the language used by newspaper writers. For example, the audience of reference varies from newspaper to newspaper, in the case of the popular press it differs from that of the quality press, as the former is interested in involving readers from lower, less-educated classes, while the latter is interested in readers from educated, high profile, "white collar" working social classes. The discussion and presentation of topics are peculiar for each news outlet and the style differs from that of other written or oral texts. In the two genres investigated, these vary accordingly, as one is generally interested in covering more local events while the other expands on internationally resonating affairs. The communicative goal enacted by the linguistic choices and style of each newspapers are peculiar of this type of genre. Despite the variation and fugacity of the language of newspapers (Hernández Longas 2001), the structural and communicative features of the language used in this type of media are strongly defined and pursued by journalists. Thus, the language of newspaper is considered in this investigation a variety of LSP. Moreover, if we define LSP against LGP (Language for General Purposes) and define the latter as the language we use in everyday interaction (Bowker, Pearson 2002), it is even more evident that the language used in newspapers is a type of LSP. Additionally, newspapers use a variety of specialized and technical terms when discussing specific issues. This aspect reflects in the corpus analysed in this study, especially considering the divide between popular and quality press.

Against this backdrop, among the features that can be connected to its language being an LSP – the corpus was divided into two sub-corpora each comprising articles from the popular and the quality press respectively - is the gross difference in size between the two sub-corpora in terms of word tokens. In fact, while the number of articles included in each sub-corpus does not differ consistently, the amount of tokens doubles in the quality press compared to the popular. The reason behind this difference results from the structural features of the two genres. The quality press is renowned for its lengthier and text-heavy content compared to the popular press which generally presents shorter articles with greater usage of multimodal elements.

Additionally, one might posit that terminology related to transgender identities is a type of LSP. Section 4 will explore this hypothesis further.

3.2. Corpus-Based Discourse Analysis (CBDA)

The existing literature on Critical Discourse Analysis (CDA) combined with Corpus Linguistics (CL) shows that the use of corpora to analyse discourses and uncover patterns of use in language has been an effective method of linguistics enquiry (see Baker 2006, 2010, Baker, McEnery 2005, Baker *et al.* 2013, Partington 2012, 2015, Venuti, Nisco 2015). On the one hand, Critical Discourse Analysis (Baker 2006; Fairclough 1992; Fairclough *et al* 2011; Hidalgo-Tenorio 2011) provides researchers with a framework that favours a more qualitative-based investigation and endorses the study of context and the socio-political implications of language use. CDA considers discourse as a form of social practice and is particularly interested both in the way in which ideology and power influence the use of language and the way language is used in society to determine specific power relations. CDA explores given linguistic phenomena to determine the way these linguistic events express specific viewpoints (Wodak, Meyer 2001) and highlights the relations between language, context, society, history and politics. On the other hand, Corpus Linguistics (McEnery, Hardie 2012; McEnery, Wilson 1996; Taylor 2008) allows for the analysis of large amounts of datasets. McEnery and Wilson (1996) explain CL as an approach for exploring language use through the analysis of a corpus and the use of a software program that performs immediate and precise calculation on the data collected. The analysis is carried out through the application of a number of tools, which allows the linguist to have a general overview of the text and identify linguistic patterns (Baker *et al.* 2013). The advantages of using corpus linguistics are rooted in the possibility to investigate extensive amounts of data at once, thus enabling the scholar to produce highly representative results.

Baker *et al.* (2008) suggest a systematic top-down framework of analysis combining CL and CDA. This moves from qualitative to quantitative analysis and back: Corpus-based Discourse Analysis (CBDA). These two methods combined together are particularly relevant for the kind of analysis being conducted in this study. In fact, computerized analysis helps to highlight important elements of the text, which undergo a deeper examination facilitating the creation of further hypothesis and a better and wider investigation of the text (Baker *et al.* 2013). Historically speaking, CDA focuses mainly on single texts or small collections of texts and is primarily based on qualitative analysis. The introduction of Corpus Linguistics has made quantitative analyses possible and consequently the identification of language patterns which unveil hidden meanings of lexical items. The main focus of CDA, i.e. to identify ideological patterns and power relations expressed through language in a text, is facilitated by the contemporaneous investigation of a plurality of texts and of large corpora.

The CBDA approach comprises a nine-step process. The analysis usually begins with a context-based investigation of the topic from an historical, political, cultural and etymological point of view. This initial evaluation enables the creation of research questions and allows the researcher to set out the parameters for the building of the corpus. The corpus, then built, is ready to be investigated by means of frequencies, keywords, collocations and other corpus tools. These procedures highlight potential sites of interest in the corpus and uncover possible discourses. Then, as if applying a magnifying glass, the analysis focuses on a small but representative set of data taken from the overall corpus, which may lead to new findings and possibly to the formulation of new hypotheses or research questions. The new hypotheses are tested through further corpus analysis, including the search for occurrences of intertextuality or interdiscursivity. This further investigation may lead to the formulation of new and final hypotheses. All final results and findings are tested through a last revision of the data. The following section offers an overview of the corpus under investigation: the TransCor.

3.2. The TransCor

The TransCor is a corpus of news articles collected from the British press in the time span stretching from January 2013 to December 2015. All articles in the corpus deal with the topic of transgender identities.

The articles were downloaded from the online platform LexisNexis,[3] an electronic database containing legal and journalistic documents through the use of the following search words: *transgender*, *transsexual*, *transvestite*, *trans*, *trannie*, *cross-dresser*, *sex change*, *shemale*, *gender reassignment*, and *dysphoria* (Zottola 2018a). The terms were chosen following a pilot study conducted on the first six months of 2015. The initial search-word list was longer than the one presented here. It contained all the search terms used by Baker (2014) in his study, plus other terms that had emerged during a background analysis of the literature related to transgender studies. The pilot study highlighted the search terms that would produce results; the other terms were eliminated.

The TransCor includes articles from eight national British newspapers: *The Guardian*, the *i*, *The Daily Telegraph* and *The Times* as representative of

[3] http://academic.lexisnexis.eu/

the quality press, and the *Daily Express*, the *Daily Mail*, the *Daily Mirror* and *The Sun* as representative of the popular press.

The corpus comprises 3,138 articles, for a total of 2,201,225 tokens. For the purpose of this analysis the TransCor is divided into two sub-corpora: the PopCor (Popular Corpus; this sub-corpus is made of all articles belonging to the popular press) and the QualCor (Quality Corpus; this sub-corpus is made of all articles belonging to the quality press). The two sub-corpora are respectively made of 1,488,352 words and 712,873 words. The following table illustrates in more detail the distribution of articles and word tokens in the two sub-corpora.

	2013		**2014**		**2015**	
	Number of articles	**Word Tokens**	**Number of articles**	**Word Tokens**	**Number of articles**	**Word Tokens**
QualCor	527	383,846	477	360,750	919	743,756
PopCor	358	197,387	424	252,943	433	262,543
Total	885	581,233	901	613,693	1352	1,006,299

Table 1
Distribution of articles and word tokens across sub-corpora and years.

One of the issues concerned with the selection of articles, for the purpose of this study, was the decision to include different genres of news articles. All of the different types of articles (news reports, editorials, opinion columns, sports news) are kept, except for the weekly television, radio and theatre schedules, since they do not add any information in terms of the discourses regarding transgender identities. In fact, I believe that each article adheres in a way to the point of view of the newspaper and therefore contributes to the externalization of the newspaper's stance on given issues. The software used for the analysis is AntConc (Anthony 2014).

4. Analyzing discourses in the British press

The analysis of the TransCor reveals a number of elements with regards to the way in which transgender people are represented as social actors in the British press and more generally speaking how language is being currently used by newspapers in this regard. Following a context-based exploration and given the purpose of this study, the focus will be on the following research questions:

- Can terminology related to transgender identities be considered a type of LSP?
- Does the representation of transgender people as social actors given by the British press thus far present considerable differences regarding the popular vs. quality dichotomy?

As it is well known among the transgender community, and as the guidelines consulted in the writing of this work specify,[4] the terms transgender and transsexual are and must be used as adjectives. Regarding this aspect, the TransCor reveals an interesting pattern. In fact, considering 10% of the overall occurrences[5] of each term in the two sub-corpora, 3.8% of the term transgender*[6] is found to occur in the QualCor as a noun, while it displays this usage in the PopCor 8.5% of the times. Trans functions as a noun in the QualCor at 7% and 0% in the PopCor; transsexual* functions as a noun 50% of the times in the QualCor and 38% of the times in the PopCor.

Other lexical choices considered in the analysis of the TransCor suggest that transgender-related terminology can be included within the definition of LSP. These patterns, mostly found in the popular press, are exemplified in the following extracts:

(1) **TRANNY** FURY AT BRIDE-SHOP BAN
A **CROSSDRESSER** was left fuming after being turfed out of a bridal shop for asking to try on a wedding dress. (*The Sun*, 12 February 2015)

(2) In another, unrelated, incident it was reported that professional drag queens have been banned from taking part in Glasgow's Gay Pride Rally because they may cause offence to those who are genuinely **transgendered**/**transsexual**/**transvestite** etc. (*Daily Mail*, 24 July 2015)

In support of the hypothesis that terminology related to trans identities is a type of LSP, the above examples show instances of the use of terminology in ways that can be defined as improper. Example (1) uses the terms *tranny* and *crossdresser* as synonyms. This type of construction can be seen as problematic for a number of reasons. First of all, the words *tranny/trannies/trannie*, retrieved throughout the corpus, are considered by some transgender people as derogatory and are classified as offensive by most guidelines on the use of inclusive language in reference to trans identities. Additionally, a tranny is not a crossdresser and it is inaccurate to use the two terms as synonyms.

Example (2) shows another pattern frequently found in the PopCor, that is the tendency to represent several different trans identities together as if they were one and the same thing. This type of construction signals inaccurate use of specific vocabulary. In fact, in example (2) two different issues arise. The first issue is related to the word *transgendered*. This deverbal noun – entailing a process of nominalization through the addition of the inflectional suffix -ed

[4] For more on this: https://www.glaad.org/reference/transgender (30.08.2018).

[5] As some of the terms have very high frequencies, 10% of the overall occurrences were used as a representative sample to identify more general patterns.

[6] The (*) at the end of the word signals that the wild card function was employed in the corpus search.

– suggests the existence of a verb ‘to transgender’. In light of what has been said so far with reference to the use of *transgender* as an adjective only, this can be defined as a misuse of terminology. The second issue is related to the collective representation given in the example. Here transsexual and transgender people are associated to transvestite, which is not a gender identity, and most importantly, has no connection to transgender identity. Considering that today the use of the term c*rossdresser* is preferred to that of *transvestite*, the repeated association of the terms might generate a misunderstanding in the reader.

The last pattern endorsing the hypothesis presented here relates to what can be defined as the informative function of the press (Zottola 2018b). In fact, many instances (23%) are present in the TransCor in which the use of terminology related to transgender identities is then followed by an explanation of the meaning of the term used, in a process of popularization of the term, that is in order to make a technical term understood by the layman. This underlines the fact that LGBT+ terminology is not always straightforward in its meaning, a feature characterising LSP (Swales 1990; Gotti 2003).

The language analysed so far belongs to a discourse community which “has acquired some specific lexis” (Swales 1990, p.26) spread, known and widely used within the given discourse community but not equally well established outside of it.

4.1. Transgender people as social actors

An additional set of results emerging from the analysis of the TransCor deals with the representation of transgender people as social actors in the British press. This paper argues that the representations offered by the popular and the quality press present considerable differences. The taxonomy proposed by van Leeuwen (1996) on social actor representation, which serves to investigate transgender people as agents/patients in the corpus, is applied to address this point. This analytical tool falls within the framework of CDA and can be defined as a “*sociosemantic* inventory of the ways in which social actors can be represented” (van Leeuwen 1996, p. 32), both linguistically and semantically. The different categorizations presented are linked to a specific linguistic and rhetorical construction and show to be effective for the purposes of this study, as the way social actors are represented usually mirrors the social practices diffused in society. van Leeuwen (1996) identifies two major categories of representation: Exclusion and Inclusion. In line with this, a newspaper can take a stance about a social actor by talking about the social actor, by doing it in a specific way or by not talking about them at all. Inclusion and Exclusion strategies can be considered as a means through which stances are legitimized or delegitimized. For example, in the collection of the corpus investigated in this study, it is necessary to include among the search terms for the newspaper articles selection the term *sex change*, as some articles, despite their talking about transition and transgender people, never use the term *transgender*, but only refer to it in terms of the pragmatic act of gender reassignment. For the purpose of this work, it is not possible to exhaustively address all categories included in the taxonomy, thus only the ones considered in the analysis will be discussed in detail.

Among the most evident differences between the PopCor and the QualCor is the use of terms related to medical issues, such as gender reassignment, to physical and bodily features of transition, or to terms generally used in relation to trans identities but unrelated to it (i.e. cross dresser). This can be signalled by, for instance, the presence in the two sub-corpora of the search terms. The following table shows the normalized (by 10,000 words) frequency of the search terms in the TransCor.

Search term	**QualCor**	**PopCor**
*transgender**	175.50	130.45
trans	58.65	12.62
*transsexual**	21.83	51.76
*sex change**	11.42	61.72
*transvestite**	18.34	29.17
gender reassignment	11.28	22.72
dysphoria	4.90	9.11
*cross-dress**	2.95	8.41
*trann**	2.83	3.50
*shemale**	0.33	0.70

Table 2
Dispersion of search terms.

From the frequency of occurrence of the search terms we can conclude that the popular press has a tendency to prefer terms that are strongly related to the physical aspect, such as *transsexual*, or to the process of transition, using both inclusive phrases, such *as gender reassignment*, and more derogatory ones such as *sex change*. A higher frequency of terms unrelated to transgender identity but associated to it in a number of contexts, see *transvestite* or *cross-dresser*, and of derogatory language is observed in the PopCor.

A strong presence of terms related to other gender identities is found in the TransCor. In fact, one of the more frequent representational patters through which transgender people are often represented, throughout the corpus, is in relation to other gender identities considered similar. This strategy can be described as an instance of Aggregation, but also of Association (van Leeuwen 1996, p. 50). In fact, through Association social actors "rather than being represented as stable and institutionalised, [are] represented as an alliance which exists only in relation to a specific activity or set of activities" (van Leeuwen 1996, p. 50-1). Examples (3) and (4) illustrate this pattern.

(3) Marriage equality for **gay, lesbian, transgender** and **intersex** Australians is the final step in a long march, beginning in 1788, for the right in Australia for all consenting couples to legally wed. (*The Guardian*, 3 December 2018)

(4) We believe that diversity is as important in the workplace as it is in the environment,' declares the agency, which has developed a programme of "unconscious bias" awareness training' for staff, as well as the establishment of an Islamic fellowship' to complement our **lesbian, gay, bisexual and transgender** and **women**'s networks'. (Daily Mail, 12 February 2014)

This kind of representation seems to be very popular with reference to transgender identities. Baker (2014, p. 221) argues that "[t]he popularity of this practice of joining together a range of identity groups who are viewed as having minority or diverse sexualities and/or genders is sometimes signified by the acronym LGBT". This pattern can be interpreted as an attempt at inclusivity; however, as Baker notes, the typical positioning of transgender people at the end of the list may also hint at a sort of hierarchical system within the same LGBT+ group. This type of representation is found both in the PopCor and in the QualCor.

Another type of construction, mainly found in *The Sun*, and the popular press more generally speaking, can be defined from the perspective of social actors representations as a form of Determination, more specifically of Categorisation (van Leeuwen 1996 pp. 52-54). This group of strategies looks at social actors by considering the characteristics that each actor shares with others. Categorisation is further divided into Functionalisation, Appraisement, and Identification. Functionalised social actors are represented in terms of an activity or a function they display, i.e. in terms of what they do. Appraisement is realized when social actors are represented through a term which evaluates them. Identification, as opposed to Functionalisation, defines social actors in terms of what they are, rather than what they do. van Leeuwen proposes three types of Identification: Classification, Relational Identification, and Physical Identification. The first category mirrors the major social classes through which society is classified, namely age, gender, race, and education. Classification, in fact, is hard to define, as it is strictly related to the society one lives in when using this strategy, and to the class the social actors belong to, whether it is associated to gender, race or education. Relational Identification represents social actors in terms of the relationship they have with a specific other and is typically possessivised, which differentiates it from Affiliation. Physical Identification depicts the social actors in terms of the physical characteristics that make them unique. This strategy is rendered through the use of pre-modifiers that hint at personal or physical details, or that specify the subject's job. Example (5) offers a (calumnious) instance of Categorisation, or more specifically Physical Identification.

(5) You would have thought they would have been outraged that Mr Sheen, knowing he was infected with HIV, would have continued to have unprotected sex with girlfriends, hookers and **pre-op transsexuals.** (*The Sun*, 20 November 2015)

Here, we find not only the presence of a pre-modifier which is not necessary to the overall understanding of the piece of news, but also a representation which depicts these people "in terms of an identity and function they share with others" (van Leeuwen 1996, p. 52), which does not represent their identity as a whole. If we focus on this pre-modifier only, there are seven occurrences throughout the corpus in which pre-op transsexual people are identified as prostitutes or as opportunists who rely on other people to live a wealthy and easy life. Similar patterns are retrieved with pre-modifiers such as *post-op*, age or nationality, as example (6) below shows.

(6) **Brazilian hooker** Naomi and **pre-op transsexual** pal Sabrina set up their sleazy sex den inside a picturesque cottage in quiet Nairn. (*The Sun*, 9 March 2014)

This pattern is mostly found in the PopCor, where the above examples were taken from.

5. Conclusion

A number of results emerge from the analysis of the specialized corpus under investigation, the TransCor. The language analysed is that of newspapers in relation to a specific topic, transgender identities. The adjective vs. noun pattern exemplified in section 4, together with the other instances of language use illustrated in the same section, highlight a use of the terminology related to transgender identities often confused or improperly applied outside of the speech community.

When discussing the extent to which terminology related to transgender identities can be considered an LSP, a number of arguments emerge. Firstly, the very specific use of this terminology, as suggested by the guidelines, and by the specificity of the community of users, grants it the status of LSP, as it is used and understood in its primary meaning within a specific community of users (Swales 1990). This argument is based on the use of unrelated terms as synonyms, the inaccurate use of given terms, and by the informative pattern retrieved in the corpus. The vocabulary analyzed in the TransCor related to transgender identities contains a number of specialized terms that many cisgender[7] people confuse or whose meaning they are not aware of. Moreover, terms such as *transgender* have undergone a process of "de-terminologization"

[7] Those people whose gender assigned at birth and gender identity correspond. It is notable that the need to insert this footnote corroborates the hypothesis being discussed.

(Meyer, Mackintosh 2000) – typical feature of LSP terminology – and made their way into general language. That is to say that lay people are now using terms initially used specifically in the in-group transgender community or in the medical environment. Specialized knowledge is at the heart of this investigation as it emerges from the analysis presented in the previous sections. The language of newspapers combined with the vocabulary used in relation to trans identities not only suggests reconsideration for what can be defined as LSP, but strongly redefines the boundaries of what is historically known as specialized knowledge. This on the one hand suggests an ever-growing advancement of this field and of the evolution of language, while on the other highlights the need of our society to embrace and endorse this language evolution in order to keep conveying meaning accurately.

As far as the distinction between popular and quality press is concerned, the investigation presented here demonstrates a tendency by the popular press to prefer terms related to the physical aspect, or to the process of transition, using both inclusive and derogatory linguistic constructions. Additionally, the PopCor presents a higher percentage of terms unrelated to transgender identity, but associated to it.

In terms of the way in which transgender people are represented as social actors, two main categories are employed in the TransCor: Association and Physical Identification (van Leeuwen 1996, p. 50). Unfortunately, some of these patterns have been found to be negative, mainly in the PopCor. However, generally speaking, it is possible to say that the press more comprehensively needs to further understand and comprehend the use of this LSP.

The hypothesis presented in this work regarding the status of terminology related to transgender identities being a type of LSP is most likely to change in the next few years as the press is increasingly becoming, together with other forms of media, a means though which this topic is being popularized and disseminated.

Bionote: Angela Zottola holds a PhD in *Mind, Gender and Language:* Language, Linguistics and ESP from the University of Napoli Federico II (Italy). She is currently a Research Fellow in Applied Linguistics at the University of Nottingham (UK). Until June 2018 she worked as an Adjunct Lecturer in English Language and Linguistics at the University of Napoli Federico II. Among her research interests are Language, Gender and Sexuality Studies, Corpus Linguistics, Queer Linguistics, Sociolinguistics, Media Discourse, Health Humanities and AVT. In her current position as a Research Fellow she is exploring narratives written by transgender patients seeking assessment in a Gender Identity Clinic in the UK. She is currently working on a monograph to be published by the end of 2019 with Bloomsbury on the representation of transgender identities in the press in the UK and Canada.

Author's address: angela.zottola@nottingham.ac.uk

References

Adampa V. 1999, *Reporting of a Violent Crime in three Newspaper Articles. The Representation of the Female Victim and the Male Perpetrator and their Actions: A Critical News Analysis*, in "Centre for Language in Social Life Working Paper No 108". Department of Linguistics and Modern English Language, Lancaster University, Lancaster.

Allan S. 1999, *News culture*, Open University Press, Buckingham/Philadelphia.

Anthony L. 2014, AntConc (3.4.3m), Waseda University, Tokyo. http://www.laurenceanthony.net/

Baker P. 2005, *Public Discourses of Gay Men*, Routledge, Oxon/New York.

Baker P. 2006, *Using Corpora in Discourse Analysis*, Continuum, London.

Baker P. 2010, *Representations of Islam in British broadsheet and tabloid newspapers 1999-2005*, in "Language and Politics" 9 [2], pp. 310-338.

Baker P. 2014, *Bad wigs and screaming Mimis: using corpus-assisted techniques to carry out critical discourse analysis of the representation of trans people in the British Press*, in Hart C. and Piotr C. (eds), *Contemporary Critical Discourse Studies*, Bloomsbury, London, pp. 211-235.

Baker P. and McEnery T. 2005, *A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts*, in "Language and Politics" 4 [2], pp. 197-226.

Baker P., Gabrielatos C. and McEnery T. 2013, *Discourse Analysis and Media Attitudes. The Representation of Islam in the British Press*, Cambridge University Press, New York.

Baker P., Gabrielatos C., KhosraviNik M., Krzyzanowski M., McEnery T. and Wodak R. 2008, *A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press*, in "Discourse and Society" 19 [3], pp. 273-306.

Bednarek M. 2006, *Evaluation in Media Discourse. Analysis of a Newspaper Corpus*, Continuum, London/New York.

Bednarek M. and Caple H. 2012, *News Discourse*, Continuum, London/New York.

Bell A. 1991, *The Language of News Media*, Blackwell, Oxford.

Bergenholtz H. and Tarp S. 1995, *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*, John Benjamins Publishing, Amsterdam/Philadelphia.

Byerly C. and Ross K. 2006, *Women and Media. A Critical Introduction*, Malden-Blackwell Publishing, Oxford/Victoria.

Bowker L. and Pearson J. 2002, *Working with Specialized Language. A Practical Guide to Using Corpora*, Routledge, London/New York.

Butler J. 1990, *Gender Trouble: Feminism and the Subversion of Identity*, Routledge, New York.

Cabré T.M. 2003, *Terminology. Theory, Methods and Applications*, in "Terminology" 9 [2], pp. 163-199.

Carter C., Branston G. and Allan S. 1998, *News, Gender and Power*, Routledge, London/New York.

Carter R. 1988, *Front pages: lexis, style and newspaper reports*, in Mohsen G. (ed), *Registers of Written English. Situational Factors and Linguistic Features*, Pinter Publishers Limited, London.

Clarke K. 1992, *The Linguistics of Blame: Representations of Women in The Sun Reporting of Crimes of Sexual Violence*, in Toolan M. (ed), *Language, Text and Context: Essays in Stylistics*, Routledge, London, pp. 208-224.

Crystal D. and Davy D. 1969, *Investigating English Style*, Longman, London.
Conboy M. 2002, *The Press and Popular Culture*, Sage Publications, London, Thousand Oaks/New Delhi.
Conboy M. 2006, *Tabloid Britain: Constructing a Community through Language*, Routledge, London/New York.
Conboy M. 2007, *The Language of the News*, Routledge, London/New York.
Conboy M. 2010, *The Language of Newspapers: Socio-Historical Perspectives*, Continuum, London/New York.
Eco U. 2015, *Numero Zero*, Penguin, London.
Fairclough N. 1992, *Discourse and Social Change*, Polity, Cambridge.
Fairclough N. 1995, *Critical Discourse Analysis: the Critical Study of Language*, Longman, London/New York.
Fairclough N., Mulderrig J. and Wodak R. 2011, *Discourse Studies: A Multidisciplinary Introduction*, Sage Publications, London/New York.
Ferraresi A. 2018, *Who Writes the Story Matters: Transgender Identity through the Lens of Citizen Journalism*, in Balirano G. and Palusci O. (eds), *Miss Man? The Adventures of Gender*, Cambridge Scholars Publishing, Newcastle upon Tyne.
Fowler R. 1991, *Language in the News. Discourse and Ideology in the Press*, Routledge, London.
Ghadessy M. 1988, *Registers of Written English. Situational Factors and Linguistic Features*, Pinter Publishers Limited, London.
Gotti M. 2003, *Specialized Discourse. Linguistic Features and Changing Conventions*, Peter Lang, Bern.
Gouveia C.A.M. 2005, *Assumptions about Gender, Power and Opportunity: Gays and Lesbians as Discursive Subjects in a Portuguese Newspaper*, in Lazar M. (ed), *Feminist Critical Discourse Analysis. Gender, Power and Ideology in Discourse,* Palgrave Macmillan, London, pp. 229-250.
Gupta K. 2018, *Response and responsibility: Mainstream media and Lucy Meadows in a post-Leveson context*, in "Sexualities".
Henry H. 1983, Are the national newspapers polarizing? *Admap*, 19 [10], pp 484-491.
Hernández Longas E. 2001, *La armonia de diferentes terminus de especialidad en el discurso de la prensa escrita*, in Palmer J.C., Posteguillo S. and Fortaner I. (eds), *Discourse Analysis and Terminology in Languages for Specific Purposes*, Publicacions de la Universitat Jaume I, Castelló de la Plana, pp. 65-74.
Hidalgo-Tenorio E. 2011, *Critical Discourse Analysis, An overview*, in "Nordic Journal of English Studies" 10 [1], pp. 183-210.
Hidalgo-Tenorio E. and Bartley L. 2015, *Constructing Perceptions of Sexual Orientation: A Corpus-based Critical Discourse Analysis of Transitivity in the Irish Press*, in "Estudios Irlandeses" 10, pp. 14-34.
Hidalgo-Tenorio E. and Bartley L. 2016, *To be Irish, gay, and on the outside. A critical discourse analysis of the Other after the Celtic Tiger period*, in "Journal of Language and Sexuality" 5 [1], 1-36.
Jucker A. 1992, *Social Stylistics. Syntactic Variation in British Newspapers*, Mouton de Gruyter, Berlin.
Litosseliti L. 2002, *'Gead to Head': Genderd repertoires in newspaper arguments*, in Litosseliti L. and Sunderland J. (eds), *Gender Identity and Discourse Analysis*, John Benjamins, Amsterdam/Philadelphia, pp. 129-148.
McEnery T. and Hardie A. 2012, *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press, Cambridge.

McEnery T. and Wilson A. 1996, *Corpus Linguistics. An Introduction*, Edinburgh University Press, Edinburgh.
Meyer I. and Mackintosh K. 2000, *When terms move into our everyday lives: an overview of de-terminologization*, in "Terminology" 6 [1], pp. 111-138.
Morrish E. and Sauntson, H. 2007, *Camp Codes and Subtle Outing: How the British Broadsheets Press Learned Slang*, in Morrish E. and Sauntson H. (eds), *New Perspectives on Language and Sexual Identity*, Palgrave, Basingstoke.
Partington A.S. 2012, *The changing discourses on antisemitism in the UK press from 1993 to 2009: A modern-diachronic corpus-assisted discourse study*, in "Journal Of Language and Politics" 11 [1], pp. 51-76.
Partington A.S. 2015, *Corpus-assisted Comparative Case Studies of Representations of the 'Arab world*, in Baker P. and McEnery T. (eds), *Corpora and Discourse Studies: Integrating Discourse and Corpora*, Palgrave Macmillan, London, pp. 220-243.
Seymour-Ure C. 1996, *The British Press and Broadcasting, since 1945*, Blackwell Publishers, Oxford.
Swales J.M. 1990, *Genre Analysis: English in Academic and Research Settings* (13th edn., 2008), Cambridge University Press, Cambridge.
Taylor C. 2008, *What is Corpus Linguistics? What the data says*, in "ICAME Journal" 32, pp. 179-200.
van Dijk T. 1988, *News as Discourse*, Lawrence Erlbaum Associates, Hillsdale.
van Leeuwen T. 1996, *The Representation of Social Actors*, in Caldas-Coulthard C.R. and Coulthard M. (eds), *Texts and Practice*, Routledge, London, pp. 32-70.
van Leeuwen T. 2005, *Introducing Social Semiotics*, Routledge, London.
van Leeuwen T. 2008, *Discourse and Practice: New Tools for Critical Discourse Analysis*, Oxford Studies in Sociolinguistics, Oxford.
Venuti M. and Nisco M.C. 2015, *Languaging the Riots: A Corpus-based Investigation of the Rioters' Identity as Reported by the British Press*, in Balirano G. and Nisco M.C. (eds), *Languaging Diversity. Identities, Genres, Discourses*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 16-31.
Williams R. 1976, *Keywords. A Vocabulary of Culture and Society*, Oxford University Press, New York.
Wodak R. and Meyer M. 2001, *Methods of Critical Discourse Analysis*, Sage, London.
Zottola A. 2018a, *Living as a Woman. The British Press on Trans Identity*, in Balirano G. and Palusci O. (eds), *Miss Man? Languaging Gendered Bodies*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 168-189.
Zottola A. 2018b, *Transgender identity labels in the British press (2013–2015). A corpus-based discourse analysis*, in "Journal of Language and Sexuality" 7 [2]: 237-262.
Zottola A. (forthcoming), *Transgender Identities in the Press. A corpus-based discourse analysis*, Bloomsbury, London.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 481-507
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p481

SCIENTIFIC CONTROVERSIES AND POPULAR SCIENCE IN TRANSLATION
Rewriting, Transediting or Transcreation?

MARIA TERESA MUSACCHIO, VIRGINIA ZORZI[1]
UNIVERSITÀ DEGLI STUDI DI PADOVA

Abstract – Over the centuries, the circulation of scientific ideas has been granted in one or a limited number of languages. Despite the advantages of avoiding a scientific Babel, popular science is largely communicated to the public using their first language(s), and is often the result of translation from other languages – most notably English. While science may partly be communicated to the public for information, at the leading edge of research it is often popularised for its newsworthiness and/or to involve the public in debates concerning social issues or political decisions. The question addressed in this paper is how the 'news' elements in popular science are mediated in the target language and culture and to what extent processes such as rewriting, transediting and transcreation are at work. Methods and strategies for science communication are compared and contrasted using an Italian and English parallel/comparable corpus of newspaper, magazine and news agency articles reporting on the recent scientific controversy over vaccines. Corpus articles are collected using the LexisNexis database. Data are checked against a small monitor corpus of key articles collected as the controversies developed. Within corpus texts, mediating strategies are tested and issues concerning the achievement of intended effects in scientific controversy popularizations are considered. The discourse of controversies will be investigated in translation as a test case for rewriting, transediting or transcreation with an eye to different audiences, while bearing in mind that the ease of communication and circulation of ideas may have blurred cultural specificities and impacted the presentation of scientific topics to some extent.

Keywords: translation; popular science; transediting; rewriting; transcreation; corpus linguistics

1. Introduction

[1] Maria Teresa Musacchio wrote sections 2.2, 4.2 and 5; Virginia Zorzi wrote sections 1, 2.1, 3 and 4.1.

Science and technology are embedded in many aspects of modern life, so that there is an increasing need for publicly available information about research activities, results and applications. Science communication takes place in a number of contexts (academic, institutional, educational, public, etc.) where language is used to construct scientific knowledge and has specific discursive features (Halliday, Martin 1993). Substantial differences between typical instances of popular and specialised scientific communication lie in their purposes and functions. Popularisation cannot be regarded as a mere simplification of specialised research accounts (Myers 1990). One of the primary sources of scientific information in non-specialised contexts is news, whose production is affected by a number of factors such as news selection, time constraints, and embargo policies from specialised journals. In communicating science, journalists do not only need to inform the audience, but also to entertain them (Siegfried 2006; Bianucci 2008). However, when science brings up contentious issues, especially concerning ethics, politics and public health, other communicative purposes, closer to argumentation and persuasion, come into play along with information and entertainment. Scientific controversies, especially when their impact is perceived as relevant at the public level, emphasise the increasing awareness on the part of society at large – including public institutions, private companies, scientists themselves, and citizens – of the importance for scientists to communicate research activities and results outside specialised contexts. The nature of science as an international enterprise, involving collaboration among scientists from different countries, and the ease of communication and circulation of ideas through the web may have blurred cultural specificities and impacted the presentation of scientific topics as regards structure, plan and rhetoric in different languages.

Linguistic studies on the discourse of public communication of science range across and combine different perspectives. Discourse-analytical approaches were adopted to investigate the blurred boundaries between specialised and popular discourses (Myers 2003), and to identify the strategies employed in the language of popular science. These include narrative patterns (Myers 1994; Seguin 2001), reported speech (Calsamiglia, López Ferrero 2003), denominations, explanations, descriptions (Calsamiglia, van Dijk 2004), text organisation, accommodation strategies, stance (Hyland 2010), proximity (Scotto di Carlo 2014), metaphors (Williams Camus 2016), and markers of newsworthiness (Molek-Kozakowska 2017). LSP studies also contributed to the analysis of popular science language, focusing on text structure and phenomena such as repetition, vagueness, expository and colloquial styles, reader construction, and attribution practices (Garzone 2006). In surveying similar phenomena on examples from different popular sources, Gotti (2012) defined popularisation as a form of rewriting scientific research documents. Rewriting was regarded as a sort of translation because it gives rise to a text that is not completely equivalent to its original and that requires recontextualisation to make it suitable for the lay public. In multidisciplinary approaches to science communication, the idea of popular science as a kind of translation has been abandoned as it projects the idea that popularisation leads to simplified, inaccurate content. At the same time, the idea emerged that science is co-constructed by the different stakeholders involved to meet the prevailing communicative needs on contextual grounds. Most linguistic investigations dealt with examples of 'established' or 'unproblematic' science at popular level, that is, they illustrated cases when scientific knowledge follows a 'standard' communication flow from researchers to lay audiences, where it is generally accepted, as experts' authority is recognised. On the contrary, little attention has been paid to situations where scientific results and applications are at the centre of public debates.

Similar developments occurred in translation, where re-contextualisation to make texts suitably accessible to a given audience is seen as a form of refraction (Lefevere 1992). This process is not described as rewriting, but as intercultural mediation (IM), which considers the impact of cultural distance (Katan 2013) and has been explored in technical and scientific translation by Sager (1994). In translation, the re-contextualisation processes deriving from IM may be subsumed under the heading of 'constrained communication' (Chesterman 2004, pp. 10-11). In other words, a mediated text is a mixture of two textual and socio-cultural events and is located in a space that lies between the two: "Mediation regularly occurs in the popularisation of scientific research articles in newspaper reporting" (Ulrych 2015). Re-contextualisation refers to the transformation of information in mediated discourse through additions, deletions, substitutions, rearrangements, and elaborations. As part of mediation, rewriting is a metalinguistic process that reinterprets or manipulates content to serve a range of ideological motives. In translation, texts are rewritten and often 'standardised' to make them more accessible to the target audience and fit the text to the target culture expectations. Mediation generally implies trying to understand what the author(s) actually meant and identifying possible sources of language or culture interference. Besides rewriting, mediation includes other modes, such as transediting and trancreation. Originally defined by Stetting (1989, p. 374) as a form of rewriting combining translation and editing, transediting is a pragmatic translation strategy meant to smooth and improve the readability of a text (Chesterman 1997, p. 112). Forms of transediting include cleaning-up transediting or linguistic adaptation to achieve target text efficiency; situational transediting or adaptation to suit the intended function(s) of the target text; and cultural transediting or adaptation to the needs and conventions of the target culture (Stetting 1989, pp. 373-377). Transcreation is a practice going beyond translation to recast the source text in a new language while preserving the intended content (Pedersen 2014). So far, studies on the translation of popular science (cf. Garzone 2006, Olohan, Salama-Carr 2011, Byrne 2012, Olohan 2016) have focused on the communication of science to the lay public for information and entertainment. Mediation as a broad concept encompassing rewriting, transediting and transcreation strategies is used here as a useful framework to account for what occurs in translating when scientific controversies are relayed.

This study focuses on mediation in a scientific controversy, where scientific knowledge is debated among actors with different interests and priorities, and non-specialised sectors of society may have a key role in its construction and development (Bucchi 2008). A cross-cultural element is also considered, in that methods and strategies for science communication are compared and contrasted using English and Italian comparable corpora of newspaper, magazine and news agency articles with a view to drawing information and analysing a parallel corpus of articles translated from English into Italian. In particular, this paper will focus on the controversies over vaccines and vaccine safety, where the plural form ‘controversies’ is used because the debate covers a variety of issues and episodes. A recent study published in *Ebiomedicine* (Larson *et al.* 2016) shows that vaccines are perceived differently across cultures and countries, which means that the debate is, at least partly, culturally defined. This has implications for the mediating strategies which might be employed in translating such controversial content from source to target language. These strategies will be considered here, along with issues concerning the achievement of intended effects on the part of authors. However, the public communication of scientific controversies has discursive and translational features that have not been fully explored. They will be investigated here as a test case for mediation, with an eye to their broader cultural context, and the presence of different audiences.

2. Method and corpus

2.1 Comparable corpora

Our method to analyse the media coverage of vaccine-related debates consists of two main steps. The first is a quantitative and qualitative analysis on comparable English and Italian corpora, whose results are used as a basis for the second step, which consists in a qualitative analysis of English-to-Italian translations of news items concerning the debate. The quantitative analysis on comparable corpora was performed on a collection of newspaper articles downloaded through Lexinexis Academic using two separate Boolean queries including the concepts of 'vaccine', 'safety' and 'risk' in the two languages.

The English texts retrieved come from a wide range of national and local newspapers mostly based in English-speaking countries – mainly the UK, the US, Canada and Australia – but also include international editions of newspapers from other countries, such as China and India. As for the Italian texts, they were published in national newspapers mainly featuring *La Stampa*, *Il Corriere della Sera*, and in local sources from the 'Quotidiano Nazionale' group (mainly *Il resto del Carlino, Il Giorno, La Nazione*). All articles were published between 2012 and 2017 – a period when the debate was lively and characterised, in some countries, by a drop in vaccination rates and an increase in cases of infectious diseases such as measles. After retrieval, the corpus was cleared of unwanted text strings (such as Lexisnexis metadata) and split into single-text files before analysis with the Antconc software (Anthony 2017). Corpus size and components are shown in Table 1.

After pre-processing, frequency lists were produced. Subsequently, keywords were extracted by comparing the retrieved texts to reference corpora of general web language (Baroni *et al.* 2009). These lists contributed to an overview of the main themes and most relevant concepts characterising the corpora. Once themes were identified, key words/phrases were further inspected through collocation analysis.

Language	**Time period**	**No. of Tokens**	**No. of texts**
English	2012-2013	106,149	250
	2014-2015	133,301	250
	2016-2017	169,434	325
Total		**408,884**	**825**
Italian	2012-2013	164,229	329
	2014-2015	129,620	324
	2016-2017	86,215	201
Total		**380,064**	**854**

Table 1
Size and components of the comparable corpora retrieved from Lexisnexis.

The qualitative analysis was performed on a smaller corpus, manually retrieved from newspaper, news agency and magazine websites (see Table 2). Here, the main focus was the exploration of metaphorical language, with particular attention to texts referring to the same piece of news reported across different sources and languages (marked in bold in Table 2). Other features surveyed included terminology, its management, and the ways in which the connections between vaccines and their side effects were dealt with.

Title				
English	**Year**	**Source**	**Tokens**	**Total**
Large study finds no vaccine link to nerve disorder	2013	*Reuters*	709	
Study Finds No Vaccine Link to Guillain-Barré	2013	*New York Times*	234	
MMR vaccine not linked to autism, even in high-risk kids	**2015**	***Reuters***	**571**	
Large Study Finds MMR Vaccine Doesn't Cause Autism, And May Lower Autism Risk	**2015**	***Forbes***	**668**	
No link between MMR and autism, major study concludes	2015	*Guardian*	649	
Vaccine refusal tied to increased risk of measles and pertussis	2016	*Reuters*	603	3,434
Italian				
Nessun legame tra il trivalente e l'autismo	2015	*Repubblica*	233	
Vaccini e autismo, un nuovo studio smentisce correlazione	**2015**	***Repubblica***	**882**	
Vaccini e autismo, nuovo studio Usa smentisce relazione	**2015**	***La Stampa***	**454**	
Salute, governo rende obbligatori 12 vaccini per asili e materna	2017	*Reuters*	165	
Vaccini, Senato approva decreto, va alla Camera	2017	*Reuters*	120	
Decreto vaccini, via libera dal Senato	2017	*Corriere della Sera*	1,165	
Vaccini, sì al decreto su obbligo a scuola: «Saranno 12 quelli richiesti»	2017	*Corriere della Sera*	710	
Vaccini, Senato approva decreto: sanzioni abbassate, monodose, niente farmacie, obbligatori da 12 a 10. Ecco cosa cambia	2017	*Il Fatto Quotidiano*	1,238	
Vaccini, sì a decreto per obbligo a scuola. Da 4 a 12 quelli obbligatori. Chi rifiuta rischia maxi multe e potestà genitoriale	2017	*Il Fatto Quotidiano*	1,171	6,138
Overall total				9,572

Table 2
Composition of the comparable corpora used for qualitative analysis.

2.2 Parallel corpora

Similarly to the comparable corpora, our parallel corpus was analysed in two steps. First, a quantitative analysis was conducted to establish similarity of the English source texts with the comparable English corpus. Our parallel minicorpus was retrieved from two sources of translations: the *Project Syndicate* website, where all articles are written in English, but translations in many languages, sometimes including Italian, are present; and *Internazionale*, an Italian weekly magazine mainly consisting of articles translated into Italian. Once the articles were identified, the English source text was retrieved as well. The *Project Syndicate* articles were published between 2013 and 2017 while the *Internazionale* articles all belonged to the same issue as they were translations of three articles from the interspecialist journal *Science*. Corpus size and composition are summarised in Table 3.[2]

Title				
English	**Date**	**Source**	**Tokens**	**Overall Total**
The Full Value of Childhood Vaccines by S. Berkley (FVCV)	18.04.2013	*Project Syndicate*	763	
An Advocacy of Dunces by D. Ropeik (AD)	16.04.2014	*Project Syndicate*	906	
Militant Islamism and Vaccine Skepticism by J. Kennedy and D. Michailidou (MIVS)	22.10.2015	*Project Syndicate*	877	
When Populism Can Kill by D. Michailidou and J. Kennedy (WPCK)	20.07.2017	*Project Syndicate*	905	
The Science of Persuasion by K. Kupferschmidt (SP)	20.04.2017	*Science*	3,511	
Vaccines on Trial by M. Wadman (VT)	20.04.2017	*Science*	2,241	
False: Vaccination Can Cause Autism/Mercury in Vaccines Acts as a Neurotoxin/Countering Mercury from Vaccines Can Make Children Better/Spreading out Vaccines Can Be Safer for Kids by L. Wessel (F)	20.04.2017	*Science*	1,188	10,391
Italian				
Il vero valore della vaccinazione infantile by S. Berkley	18.04.2013	*Project Syndicate*	801	
Quando attivismo fa rima con egoismo by D. Ropeik	16.04.2014	*Project Syndicate*	1,031	
L'islamismo militante e lo scetticismo sui vaccini by J. Kennedy and D. Michailidou	22.10.2015	*Project Syndicate*	928	
Quando il populismo rischia di uccidere by D. Michailidou and J. Kennedy	20.07.2017	*Project Syndicate*	990	
La guerra dei vaccini by K. Kuperferschmidt	12.05.2017	*Internazionale*	3,570	
Parola ai giudici by M. Wadman	12.05.2017	*Internazionale*	2,267	
I miti da sfatare by L. Wessel	12.05.2017	*Internazionale*	1,361	10,948

Table 3
Size and composition of the parallel corpus.

Once frequency lists were produced the ten most frequent words were compared with keywords in the English comparable corpus to establish whether similarity was enough to warrant comparison and contrast. On the other hand, the most frequent words in the Italian texts in the parallel corpus were compared with those of the English source texts to check whether they 'leant to' their English source texts as is generally assumed of translations.

Second, qualitative analysis was performed on the whole parallel corpus, but focus was on the *Internazionale* translations as the texts were meant for a readership of scientists as source texts, but were translated into Italian for the lay public. The hypothesis to be tested in this case was whether the change in readership caused more additions, deletions, substitutions, rearrangements, and elaborations in translation and whether these could be described as forms of rewriting, transediting or transcreation.

[2] Articles are followed by an acronym in brackets to be easily identifiable when examples are quoted in Section 4.2.

3. Quantitative analysis of the comparable corpora: results and discussion

3.1 Comparable corpora

Quantitative analysis relies on established tools in corpus-related approaches, starting with an overview of frequency lists, obtained by excluding a set of stopwords and applying lemmatisation.[3] The lemmas found at the top of such lists were used as the main reference for a more in-depth analysis of the single word forms contributing to each lemma and potentially revealing distinct patterns of use and shades of meaning. Subsequently, keywords were extracted. Keyness is a quality of words whose "frequency (or infrequency) in a text or corpus is statistically significant, when compared to the standards set by a reference corpus" (Bondi 2010, p. 3). Scott, Tribble (2006, p. 56) see keywords as reflecting the 'aboutness' and style of a text. Keyness can also point to authors' stance and identity, and to assumptions, values and beliefs of the discourse community in which a text was produced. If combined with contextual information, keywords can help identify the conceptual and organisational structures of texts beyond single words (Bondi 2010; Scott 1997; Scott, Tribble 2006). Moreover, analysing keyness in two different corpora dealing with a similar subject can provide empirical evidence about how topic representation in different contexts (Stubbs 2010). This idea is applied here to comparable corpora, adopting a cross-linguistic and cross-cultural perspective.

The core themes emerging from frequency (Table 4) partly reflect the initial query made to Lexisnexis.

	English		Italian	
Rank	**Lemma**	**Relative Frequency**	**Lemma**	**Relative Frequency**
1	*VACCINE*	1.34%	*VACCINO*	0.70%
2	*RISK*	0.75%	*ANNO*	0.52%
3	*FLU*	0.68%	*VACCINAZIONE*	0.52%
4	*CHILD*	0.63%	*RISCHIO*	0.45%
5	*SAY*	0.62%	*POTERE*	0.44%
6	*HEALTH*	0.57%	*CASO*	0.32%
7	*CAN*	0.49%	*MALATTIA*	0.31%
8	*VACCINATION*	0.44%	*TUTTO*	0.26%
9	*YEAR*	0.43%	*BAMBINO*	0.24%
10	*DISEASE*	0.42%	*MEDICO*	0.24%

Table 4
Frequency lists in the English and Italian comparable corpora showing top-ranking concepts as lemmas.

[3] Someya's lemmatisation list, retrieved from http://www.laurenceanthony.net/software/antconc/, was applied to English, while Měchura's lemmatisation list, retrievable from http://www.lexiconista.com/datasets/lemmatization/, was applied to Italian.

Vaccines and health risks occupy top ranks; on the other hand, the idea of safety, also part of the query, is not so prominent. This suggests that popular press media tend to frame the controversy in terms of risks and dangers – whether they are associated to vaccines or vaccine refusal – rather than emphasizing safety.

Together with most frequent words, collocations are a major component in corpus-based studies and are here extracted and analysed on a set of items selected from frequency and keyword lists, which were regarded as particularly informative for this analysis. According to Sinclair (1991, p. 170), collocation can be defined as "the occurrence of two or more words within a short space of each other in a text": it is therefore a type of syntagmatic relation between words (Stubbs 1996). With reference to statistics, Baker *et al.* (2008, p. 278) define it as "the above-chance frequent co-occurrence of two words within a pre-determined span." Since collocates of a word contribute to its meaning and function, collocations can highlight different meanings of the same word form, its dominant phraseology and its own semantic field (Hunston 2002). A concept closely related to collocation is that of semantic prosody, through which the co-text of a word can reveal information about the connotation(s) it carries in the analysed texts (Stubbs 1996; Sinclair 2004).

Rank	**Word**	**Relative Frequency**	**Keyness**	**Collocates**
1	*vaccine*	0.98%	9636.35	*tdap,pentavalent, hesitancy, attributed, inactivated, attenuated*
2	*flu*	0.68%	6739.88	*swine, bird, seasonal, beat, unpleasant, season*
3	*risk*	0.64%	4984.27	*perception, minimise, poses, contracting, increased, posed*
4	*vaccination*	0.44%	4351.09	*Kingston, movement, mass, clinic, rates, programs*
5	*vaccines*	0.36%	3572.14	*clumping, improperly, expired, ineffective, stored, Novartis*
6	*health*	0.57%	3194.84	*institutes, organization, digest, chief, organisation,cabinet*
7	*date*	0.03%	2573.06	*immunisations, up, vaccinations,their, with, to*
8	*vaccinated*	0.23%	2277.8	*fully, against, getting, get, start, mothers*
9	*said*	0.51%	2228.53	*Byrne, Sheppeard, Lauzon, Perrett, spokeswoman, ms.*
10	*disease*	0.28%	2172.75	*obstructive, pulmonary, kidney, emphysema, centers, neurological*

Table 5
English keyword list with collocations.

Both *VACCINE/VACCINO*[4] (mostly indicating the substance) and *VACCINATION*/*VACCINAZIONE* (mostly indicating the treatment) are among the core items in the analysed corpora (cf. Tables 5 and 6). Collocations suggest that in English, the information provided about vaccines is mainly classificatory, i.e. identifying different types of vaccines – often pre-modifiers such as *tdap,*[5] *pentavalent, inactivated, attenuated, bcg*.[6] In English, *vaccines* also collocates with *link* and *autism*. Concordance inspection revealed that mentions of the vaccine-autism link is often preceded by negations or verbs such as *debunk*, in line with the scientific consensus about the safety and importance of vaccines. Groups of people who reject such consensus are referred to as ***anti****-vaccination* ***movement***, as emerges from the collocates of *vaccination*. Synonyms found in the corpus are *anti-vaccine*, *anti-vax*, and *anti vaxxers*. Corresponding actors in Italian do not strongly collocate with *VACCINO/VACCINAZIONE*; they were nevertheless identified as *anti-vaccino/i, anti-vaccinazione/i, anti-vaccinatori, no vax*.

Classificatory information also appears in the collocates of the Italian *vaccino* (*coniugato, tetravalente, antipneumococco*), and, in part, *vaccinazione* (*antipneumococcica*); the 'public service' dimension is stronger than in English, and spreads across all word forms. This dimension includes:

- instructions for those who need to be vaccinated, as in *Chiunque può* ***acquistare*** *il* ***vaccino*** *in farmacia* ('anyone can purchase the vaccine at the pharmacy');
- public recommendations, as in *La* ***vaccinazione*** *viene* ***raccomandata*** *anche per categorie di soggetti che [...]* ('vaccination is as well recommended to groups of people who [...]'); and
- legislative aspects, as in *Quali sono le* ***vaccinazioni obbligatorie****?* (Which vaccinations are compulsory?).

As emerges from the analysis, vaccination is not only to do with scientific facts, but also with political decisions and with people's everyday life. Particularly in the Italian corpus, vaccines are represented as a service that can be obtained in various ways by citizens who need it. In some cases it is even explicitly referred to as something legally mandatory. This might reflect a tendency of the Italian culture to deal more explicitly with power, rules and authority, and to nominally attribute them greater relevance than in the UK and other English speaking countries (De Mooij 2004).

[4] Here, 'word form' (as equivalent to wordform and word-form) is used to indicate the different forms subsumed under the same lemma (see Sinclair 1991, p. 41).
[5] Tetanus, diphtheria and pertussis.
[6] Bacillus Calmette-Guérin, primarily used against tuberculosis.

Italian				
Rank	**Word**	**Rel. Freq.**	**Keyness**	**Collocates**
1	*vaccino*	0.36%	3348.01	*coniugato, tetravalente, antipneumococco, acquistare, adiuvato, offerto*
2	*vaccini*	0.34%	3120.42	*obbligatori, antinfluenzali, somministrati, sicuri, blocco, disponibili*
3	*vaccinazione*	0.31%	2865.49	*antipneumococcica, consigliata, raccomandata, gratuita, straordinaria, obbligatoria*
4	*rischio*	0.38%	2720.72	*Reye, appartenenti, categorie, considerate, fattore, rientrano*
5	*vaccinazioni*	0.21%	1896.71	*eseguite, obbligatorie, effettuate, crollo, raccomandate, calo*
6	*influenza*	0.21%	1734.84	*stagionale, complicanze, arrivo, prevenire, vera, sue*
7	*medici*	0.22%	1672.63	*infermieri, federazione, famiglia, fimmg, sindacato, pediatri*
8	*malattie*	0.20%	1565.04	*emopoietici, metaboliche, renali, malassorbimento, dismetaboliche, trasmissibili*
9	*virus*	0.17%	1486.67	*papilloma, costituire, umani, influenzali, batteri, fonte*
10	*asl*	0.13%	1086.03	*Brescia, ambulatori, igiene, Milano, Toscana, dipartimento*

Table 6
Italian keyword list with collocations.

Risk is another major concept in both corpora. The English form *risk*, which most contributes to the frequency of the lemma *RISK*, is predominantly a noun.[7] Concordances and collocates of *risk* and *risks* suggest an association to general health problems rather than vaccines (e.g. *[...] to* ***minimise*** *the* ***risk*** *of* ***contracting*** *the virus*; *The unvaccinated child* ***poses*** *no* ***risk*** *to vaccinated children*; *Every child would be at* ***risk*** *of* ***contracting*** *a preventable disease*), which is coherent with the previously observed negation of the vaccine-autism link. Accordingly, the strong collocation between *risk* and *perception* indicates that *risk perception* (perhaps as opposed to real risk) is mostly discussed in relation to misinformed concerns and reservations about vaccines. A similar view is expressed through the collocates *outweigh* and *benefits*, realising the proposition that the benefits of vaccines outweigh its risks. On the other hand, collocates such as *increased* and *greater* are used to inform about factors which may enhance particular health risks (e.g. *Pregnant women with influenza have an* ***increased risk*** *of complications*). All these observed patterns of use may partly perform persuasive functions, which are particularly important in the argumentation that is typical of scientific controversies. Collocations of the Italian *rischio* were instead affected by recurrent strings of texts identifying at-risk categories for

[7] *Risk* concordance lines were tagged using the CLAWS online POS tagger. Out of 2607 occurrences, 2026 (77.7% of the total occurrences) were tagged as nouns and 217 (8.3%) were tagged as adjectives as part of pre-modifying forms such as *high-risk areas*. The remaining forms (14.0%) were either tagged as verbal forms or remained untagged.

seasonal flu, with the purpose of advising them on vaccine availability and procedures, as in the following examples: *La vaccinazione è gratuita [...] per le seguenti categorie: [...] bambini e adolescenti in trattamento a lungo termine con acido acetilsalicilico, a* ***rischio*** *di Sindrome di* ***Reye*** *in caso di infezione influenzale* ('Vaccination is free of charge [...] for the following categories: [...] children and young people under long-term treatment with acetylsalicylic acid, who risk developing Reye syndrome in case of flu infection'); *Nelle* ***categorie*** *a* ***rischio*** *rientrano [...]* ('At-risk categories include [...]').

Among keywords, influenza (*flu, influenza*) is the single vaccine-preventable disease that stands out in both languages, due to the widespread annual flu vaccination campaigns reported in newspapers. In general, diseases (*disease/malattia*) are also central in both corpora for their semantic relation to vaccines: they tend to collocate with classificatory modifiers, and can be medical terms. English terms are sometimes partly defined or exemplified, as in: *chronic* ***respiratory disease****, including chronic* ***obstructive pulmonary disease****, bronchitis or* ***emphysema***, or in *people with [...]****neurological disease****, such as Parkinson's or Motor Neuron Disease*. Often, they are left undefined, as part of lists of conditions that may cause severe complications in case of flu, e.g. *those with chronic* ***respiratory disease****; chronic heart disease; chronic kidney disease; chronic liver disease; chronic* ***neurological disease****; diabetes and immuno-suppression*. In Italian, where the plural form *malattie* ('diseases') is predominant, term collocates are mostly included in similar lists, and are left undefined, e.g. *malattie croniche a carico dell'apparato respiratorio, circolatorio, renale,* ***malattie*** *degli organi* ***emopoietici,*** *diabete* ('chronic respiratory, cardiovascular or kidney diseases, hematopoietic organ diseases, diabetes'). Overall, concordances reveal that Italian texts tend to provide less explanatory and definitional material, maybe taking that knowledge for granted for readers included in at-risk groups. This confirms that Italy is a high-context culture where much information remains implicit, as opposed to English-speaking countries, which are lower-context cultures where information is made explicit (Katan 2003, p. 183) and provides useful guidelines for translation for the English-Italian language pair.

Reference to experts' statements and opinion is key in science communication, even more so during scientific controversies. Here, the role of experts/mediators between scientific notions and the public is attributed to health care professionals and, less frequently, to scientists. There is a range of noun phrases to identify actors in the medical professions; some of them have been compared between the two corpora. It was observed that Italian texts emphasise the centrality of GPs (*medico*/plural *medici*), usually further specified by post-modifiers *di base*, *di famiglia, curante, di medicina* *generale*, which all converge towards the figure of 'family doctor'/GP. *Dottore* ('doctor') and *pediatra* ('paediatrician') are used with a similar meaning. Collocations of *medico* and *medici*, and the high keyness of *medici*, point to their role as health care – including vaccination – providers, but also as trusted people who can be consulted for useful information and advice. An example is *"i soggetti appartenenti alle categorie a rischio potranno rivolgersi al proprio* ***medico di famiglia****"* ('people in at-risk categories can consult/visit their family doctors'). Moreover, collocates of the plural form identify doctors as a homogeneous category through reference to several associations and by sometimes including them among at-risk categories. English has corresponding forms, mainly *GP/general practitioner*. The local and personal element can partly be found in the use of *doctor* and *physician* – especially when preceded by possessive determiners and/or the noun pre-modifier *family*. However, this shade of meaning is less frequent in English and was not found among key items. The importance of health care services and their workers with respect to this controversy somehow moves representatives of the scientific community to a more peripheral role. Researchers and scientists are less frequent than doctors in both languages, although in the English corpus they are almost three times as frequent as in the Italian one, perhaps reflecting a slightly higher reliance on their statements and opinions in these texts.

The occurrences of *health care worker(s)/practitioner(s)/professional(s)*, etc. contributed to the overall frequency of the lemma *HEALTH*, at the top of the English frequency list. However, most of its strongest collocates – *chiefs, institutes, organisations, cabinet, ministry* – indicate that it is mostly intended as part of institutional organisations and procedures. The fact that there is no direct, equally frequent item in Italian may be due to a more lexically fragmented way to refer to heath care institutions; but it could also indicate a slightly lower level of trust and consideration towards public institutions in general.

3.2 Parallel corpora

The core themes emerging in the comparable corpora are largely confirmed in the English component of the parallel corpus (Table 7). Albeit with different frequencies, probably due to different preferences in creating cohesion,[8] Italian translations appear to be quite close to the English source texts, at least for the five most frequent items. The higher number of tokens (10,948) of the Italian component compared to the English one (10,391) is in line with the general rule that translations are longer than their originals. This

[8] Unlike English, Italian shows a preference for cohesion created with devices other than lexical repetition wherever possible.

also indicates that changes in translation include a lower number of deletions. Another element that signals the translation 'status' of the Italian articles and marks their deviation from English articles is sentence length: at an average of 25.25 words, it is not only higher than the English source texts (22.01 words) but also reflects the current standard sentence length in the Italian press (Bonomi 2002).

English			Italian	
Rank	**Keyword**	**Relative Frequency %**	**Keyword**	**Relative Frequency %**
1	*VACCINE*	1.35%	*VACCINO*	1.30%
2	*CHILD*	0.68%	*POTERE*	0.58%
3	*VACCINATION*	0,64%	*BAMBINO*	0.49%
4	*PARENT*	0.56%	*VACCINAZIONE*	0.47%
5	*SAY*	0.55%	*GENITORE*	0.43%
6	*CAN*	0.49%	*ANNO*	0.35%
7	*GOVERNMENT*	0.38%	*MALATTIA*	0.34%
8	*HEALTH*	0.38%	*VACCINARE*	0.34%
9	*INJURY*	0.35%	*CASO*	0.31%
10	*ONE*	0.34%	*SE*	0.29%

Table 7
Frequency lists in the English and Italian parallel corpora showing top-ranking concepts.

Calculation of keyness using the Lexinexis comparable corpora as reference corpora indicates that English source texts are more argumentative/persuasive as evidenced by the comparatively high frequency of words such as *say, opposition, advocates, militants* and *persuasion.* Negative keyness concerns words such as *safe, risk* and *vaccine*, which suggest a discursive strategy attempting to avoid controversial points. Keyness of the Italian parallel component exhibits higher frequency of the concepts relating to society and government – *comunità, governo/i, programme, sanità* – or to argumentation – *prove* and *autismo*. Keyness of words referring to government and society can again point to the fact that the Italian texts are translations as relationship with power and authority in Italy is culturally different from what that of English-speaking countries. Moreover, the lower frequency of the Italian equivalents of health, *salute* and *sanità*, can partly be explained with the change of name of the corresponding services and ministry, once referred to as *sanità* and now termed *salute*. The keyness of *sanità* in the Italian texts could lend support to the idea of conservatism in the language of translations (Baker 1996).

Based on the analysis of collocations in the comparable corpora, the same keywords were investigated in the parallel corpus to establish to what extent they reflected the general trends in the representation of controversies over vaccines. As Table 8 below indicates, given the size of the parallel corpus, collocations are more restricted in number and the most frequent ones are not equally key in comparable corpora. Italian equivalents are shown to highlight translation strategies.

English	**Italian**
flu/measles/MMR/tetanus vaccine	vaccino antinfluenzale/contro il morbillo/trivalente/ antitetanico
childhood vaccines/value of vaccines	vaccinazione infantile/valore dei vaccini
at risk/autism risk	a rischio/rischio di autismo
polio vaccination/vaccination rates/rates of vaccination/vaccination programs	vaccinazione antipolio/tasso di immunizzazione/ programmi/campagne di vaccinazione
World Health Organisation//Minister of Health//public health	Organizzazione Mondiale della Sanità//ministro della salute/sanità//salute/sanità pubblica
vaccinated against	vaccinato/a/i/e contro/immunizzato/a/i/e da
Centers for Disease Control	centri per la prevenzione e il controllo di malattie

Table 8
English and Italian keyword list from our parallel corpus with collocations.

As in the comparable corpora, collocations with vaccine(s) and vaccination concern information that is mainly classificatory. For collocations that are actually terms, Italian reflects the tendency to use medical terms of Greek or Latin origin where English draws on general language: *MMR vaccine* → vaccino trivalente, *vaccinated against* → vaccinato contro but also immunizzato da. *Risk* is present in concordances that downplay the association with vaccines.

Disease is mostly used in two collocations, *infectious diseases* and *Center(s) for Disease Control and Prevention* (*CDC*). The variety of diseases found in the comparable corpora is not present here, while doctors are referred to in general terms and family doctor has only one occurrence. Physician(s) is even more infrequent and all occurrences of scientists are found in the main *Science* article and in “An advocacy of Dunces”, which in the source texts aim to help scientists understand how they can best convince people that vaccines are safe and that it is important to immunise people. Both *doctor(s)* and *physician(s)* are translated as *medico/i*. Like scientist(s), researchers is used – though less frequently – in the same two texts and rendered literally as *scienziati* or *ricercatori*. This does not consider that in Italian there is a tendency to avoid these general references and replace them with the corresponding abstract concepts, namely science and research.

The consistently lower frequency of key words or terms in Italian translations also suggests an attempt to avoid lexical repetition to create cohesion and rely on strategies such as the use of synonyms or relative clauses that are standard practice in Italian writing. Together with the qualitative analysis that follows, it contributes to creating a full picture of translation features. Contrasting information gathered from the comparable and the parallel corpora gives a clearer idea of what can be regarded as more or less standard in the source language, English, and in the target one, Italian.

This allows translators not only to draw guidelines for translation, but also to gain insights for translation revision and quality assessment.

4. Qualitative analysis: Results and Discussion

4.1 Comparable corpora

The qualitative analysis was performed on texts concerning the debate on the MMR-ASD association, as shown in the frequency lists in Table 9.

	English		**Italian**	
Rank	**Keyword**	**Relative frequency %**	**Keyword**	**Relative frequency %**
1	*VACCINE*	2.13%	*VACCINO*	1.27%
2	*CHILD*	1.69%	*VACCINAZIONE*	0.67%
3	*AUTISM*	1.11%	*ANNO*	0.65%
4	*STUDY*	0.99%	*OBBLIGATORIO*	0.57%
5	*MMR*	0.76%	*BAMBINO*	0.52%
6	*RISK*	0.76%	*DECRETO*	0.46%
7	*ASD*	0.70%	*SCUOLA*	0.46%
8	*FIND*	0.61%	*AUTISMO*	0.44%
9	*PEOPLE*	0.58%	*SANZIONE*	0.44%
10	*VACCINATE*	0.50%	*SALUTE*	0.36%

Table 9
Frequency lists of texts used for the qualitative analysis.

In particular, four of the collected texts (two in English and two in Italian) report on a 2015 study published in the *Journal of the American Medical Association*, stating that there is no causal link between the MMR vaccine and ASDs in children.

Overall, the analysed corpora helped highlight some linguistic differences, which might be grounded in culture. For example, the frequency of reference to norms and laws in Italian words such as *decreto* (decree), *obbligatorio* (mandatory) and *sanzione* (sanction, penalty, punishment) may point to cultural difference in power relantioships. These, together with recurring patterns highlighted by concordances, should be taken into consideration when approaching translation. On the other hand, there are similarities between the two languages: most importantly, they shared many core themes; moreover, they used some linguistic devices – lexical and co-textual – to perform more of a persuasive and argumentative function rather than an informational one.

4.2 Parallel corpora

For the qualitative analysis of the parallel corpora, the three main categories identified – rewriting transediting and transcreation – were divided in components and a count of occurrences was taken for both the *Project Syndicate* and the *Science* article translations. As transcreation can take many different forms, instances of transcreation were detected and then labelled as components. What emerged from analysis was that transcreation was mainly used for headlines and standfirsts, though *Science* translations in *Internazionale* also have an added box (see below). Categories and instances are summarised in Table 10. To provide a better picture of the translations, 'unsuccessful' instances – i.e. attempts at rewriting, transediting or transcreation that resulted in non-standard collocations, unnatural word order, unsuitable adaptations to the target culture etc. – are also listed, preceded by minus signs.

Category and components	*Project Syndicate* article translations		*Science* article translations	
REWRITING	**36**	**–37**	**47**	**–19**
Additions	+5	–2	+2	–1
Deletions	+6	–6	+11	–2
Substitutions	+6	–16	+11	–11
Rearrangements	+14	–13	+20	–5
Elaborations	+5	–9	+3	none
TRANSEDITING	**36**	**–47**	**46**	**–27**
Linguistic transediting	+9	–26	+4	–14
Situational transediting	+19	–13	+30	–8
Cultural transediting	+8	–8	+12	–5
TRANSCREATION	**5**	**0**	**16**	**0**
Headlines and paragraph titles	+1	–	+12	–
Standfirsts	+4	–	+3	–
Boxes	none	–	+1	–
Overall totals	**77**	**–84**	**109**	**–46**

Table 10
Successful (+) and unsuccessful (–) instances of rewriting, transediting and transcreation in *Project Syndicate* and *Science* article translations.

As can be seen, the instances of rewriting, transediting and transcreation are higher overall in the *Project Syndicate* article translations (161) compared to the *Science* ones (155) and the number is even greater in relative terms as the former articles are shorter than the latter. Yet, the number of positive cases of rewriting, transediting and transcreation is much higher in the *Science* translations (109) than in the *Project Syndicate* ones (77). In two categories out of three, unsuccessful attempts in *Science* translations published by *Internazionale* are consistently lower in number than in *Project Syndicate* ones. In both sets of translations, the most frequent successful adaptations concern rearrangements as forms of rewriting – typically use of 'natural' phrase and clause order in the target language – and situational transediting, while the highest number of failed attempts is found in substitutions as forms of rewriting and in linguistic transediting. Though overall *Project Syndicate* translations into Italian are closer to their source texts because no change in readership is implied, results suggest that translators are not completely aware of modes of argumentation in the debate over vaccines in Italy and their linguistic implications.

In the following example from *Project Syndicate*, the successful word order rearrangement (*la possibilità di eliminare malattie*) and addition of *negli ultimi tempi* at the beginning of the sentence is partly offset by the literal translation of *popular resistance* (opposizione popolare instead of ‘opposizione diffusa’) as a failed case of linguistic transediting and of *governments* (governi) in the last sentence, where cultural transediting would suggest the use of *stato* (state) as the personified institution in charge of educating citizens.

> No medical or technical obstacles are blocking us from eradicating preventable infectious diseases such as measles or polio. Rather, the biggest hurdle has been popular resistance to vaccination. By allowing parents to make uninformed decisions about the health of not just their own children, but their entire community, the Syriza government is only adding to the problem. Governments should be educating the public to improve overall coverage, not validating unfounded fears about vaccine safety. (WPCK)
> Negli ultimi tempi, la possibilità di eliminare malattie infettive prevenibili non è stata inficiata da ostacoli di natura medica o tecnica quanto da un’opposizione popolare ai vaccini. Permettendo ai genitori di prendere decisioni poco circostanziate sulla salute non solo dei propri figli ma anche della propria comunità, il governo di Syriza non fa che aggravare il problema. I governi dovrebbero educare i cittadini a migliorare la copertura generale, non avallare timori infondati sulla sicurezza dei vaccini.

In *Science* translations additions as forms of rewriting are used to guide readers in understanding the authoritativeness of the opinions reported and in making sense of the controversy over vaccines. In the following extract from the “False” articles, a specification is provided of what kind of publication *The Lancet* is, *medical license* undergoes cultural transediting and becomes *albo dei medici* with the added gloss “autorizzati ad esercitare la professione”. Rearrangement as rewriting is found in the different Italian paragraphing, in phrase order in sentence three and in syntax, which involves joining of sentences and use of standard Italian punctuation such as the replacement of a full stop with a colon in Italian. Situational transediting is clear both in the translation of *the claim began to unravel* and the deletion of *Citing further concerns about ethics and misrepresentation*, which reflect the source text interest in following standard procedures in publication, retraction and the ethics of scientific research, but are not all that relevant to the general public. Finally, term formation, frequency of use and term transparency for a lay audience are at play in the choice of using *vaccino trivalente* for *MMR vaccine*.

> In 1998, UK doctor Andrew Wakefield published a study in The Lancet suggesting that the measles, mumps, and rubella (MMR) vaccine could trigger autism. In the years after, MMR vaccination rates in England dropped below 80%. But the claim began to unravel in 2004 after journalist Brian Deer reported undisclosed conflicts of interest. Wakefield had applied for a patent on his own measles vaccine and had received money from a lawyer trying to sue companies making the MMR vaccine. Citing further concerns about ethics and misrepresentation, The Lancet retracted the paper in 2010. Shortly after, the United Kingdom's General Medical Council permanently pulled Wakefield's medical license. (F)
> Nel 1998 il medico britannico Andrew Wakefield pubblicò sulla rivista The Lancet uno studio in cui affermava che il vaccino contro il morbillo, la parotite e la rosolia (Mpr) poteva provocare l'autismo. Negli anni successivi nel Regno Unito la copertura del vaccino trivalente tra i bambini di due anni scese sotto l'80 per cento. Le affermazioni di Wakefield furono messe in discussione nel 2004, quando il giornalista Brian Deer rivelò che dietro c'erano segreti conflitti d'interesse: Wakefield aveva chiesto di brevettare un suo vaccino per il morbillo e aveva ricevuto soldi da un avvocato che stava cercando di fare causa alle aziende che producevano il vaccino trivalente.
> The Lancet ritirò l'articolo nel 2010 e poco dopo il General medical council britannico radiò Wakefield dall'albo dei medici autorizzati ad esercitare la professione.

Failure to provide necessary additions as part of rewriting can adversely affect translations. In the sentence below, *fraudulent* rendered as *tendenzioso* downplays the extent to which the population was deceived as the extract above suggests. An addition such as *articolo tendenzioso poi ritrattato* or *articolo falso e tendenzioso* would have been required to relay the argument, while the different ways of creating cohesion in English and Italian are recognised in the linguistic transediting of the initial *and* as *inoltre*:

> And many high-income countries have experienced measles outbreaks in recent years, owing to fears about vaccinations that began with the publication of a fraudulent paper in the British medical journal The Lancet in 1998. (WPCK)
> Inoltre, negli ultimi anni molti paesi ad alto reddito hanno avuto a che fare con epidemie di morbillo a causa dei timori suscitati da un articolo tendenzioso sui vaccini che uscì sulla rivista medica Britannica The Lancet nel 1998.

Elaborations, situational and cultural transediting are present in the following paragraph, where relevance of some phrases in the argumentation is clearly recognised. *Of course* is translated pragmatically as *inutile dirlo* – an instance of cultural transediting, which is also the chosen strategy to render the relative clause *what those advocate believe* with a noun phrase (*sulla base delle loro convinzioni*). Two clauses are elaborated on to translate the rhetoric of the argument effectively – *Though the evidence is clear that* > *Sebbene sia dimostrato che* and *that its opponents stubbornly claim it does* > *che i loro oppositori si ostinano a denunciare*, while the standard collocation *dangerous territory* (*terreno pericoloso*) appears a faulty case of linguistic transediting as *terreno minato* is much more frequent in Italian in similar contexts.

> This is, of course, dangerous territory. Though the evidence is clear that vaccination does not cause the harms that its opponents stubbornly claim it does, any effort by a government to restrict speech is worrying. No free society should permit its government to decide which advocacy groups can say what, based on what those advocates believe. (AD)
> Qui siamo, inutile dirlo, su un terreno pericoloso. Sebbene sia dimostrato che i vaccini non provocano i danni che i loro oppositori si ostinano a denunciare, qualsiasi tentativo da parte di un governo di limitare la libertà di espressione è preoccupante. Nessuna società libera dovrebbe consentire al proprio governo di decidere quali gruppi di pressione possano parlare, e cosa debbano dire, sulla base delle loro convinzioni.

Situational transediting requires translators to identify the role phrases or clauses play in context with a view to finding good functional equivalents in the target language. In the following example, the functions of informal language – *only the evidence showed/and that by spreading their fears/*– and *you* – and direct questions are recognised and suitably relayed into Italian:

> Imagine that a group of advocates tried to alert the public to a danger that they perceived, only the evidence showed that the danger was not real, and that by spreading their fears, this group was causing people to behave in ways that put the wider public – and *you* – at risk. What would you do? What should the government do? (AD)
> Immaginate se un gruppo di attivisti tentasse di mettere in guardia la gente da un pericolo da loro percepito, e poi venisse fuori che non solo tale pericolo non era reale, ma che dando voce alle proprie paure questo gruppo ha spinto altre persone verso comportamenti potenzialmente dannosi per l'intera comunità, di cui voi stessi fate parte. Cosa fareste? E cosa dovrebbe fare il governo?

Strategies are also used to replace informal references to medical concepts in English with the corresponding term in Italian as in *after that tetanus shot* → dopo l'incidente con l'antitetanica, *vaccination effort* → la campagna di vaccinazione, *long-term neurodevelopmental damage* → disturbi dello sviluppo neurologico a lungo termine. In some cases, substitutions are not equally effective, as in the following example, where the most frequent equivalent of *eradication* is *eradicazione* while it is the verb *eradicate* whose most frequent equivalent is *sradicare*. The example also includes a case of informal language – *What is standing in the way* – which is replaced by a burocratic phrase (*elementi ostativi*) instead of a more formal equivalent:

> What is standing in the way of the virus's eradication is not medical or technical constraints, but political resistance to the vaccination effort. (WPCK)
> Gli elementi ostativi allo sradicamento del virus non sono i vincoli medici o tecnici, bensì le resistenze politiche rispetto a una campagna di vaccinazione.

In *Internazionale* translations from *Science* situational transediting is most evident where the academic setting familiar to the readers of the *Science* articles has to be evoked for the lay public of the *Internazionale* target texts. In the following example, *Plos One* is quite familiar to the readers of the source text, but needs to be glossed in translation. Similarly, the typical way of arguing in science by highlighting method and data – 'a strong preditor of antivaccine sentiments', 'little evidence of a link' – requires situational transediting to make it easily understandable for the target readers. Adaptation to the intended function is sometimes achieved at the expense of target text fluency, as is clear in the sentence starting with 'Mentre, nonostante (…)' below, which is a clear example of failed rearrangement, though the sentence was not left in brackets as in the source text to avoid introducing punctuation that is still non-standard in Italian. Finally, a substitution that makes it more difficult to follow the argument in Italian is the translation of *libertarian parents* as *sostenitori del libero mercato* where it is clear that in the source text reference is made to parents who believe in freedom as a core principle, which may or may not have something to do with a free market.

> In a study published in *PLOS ONE*, Lewandowsky reported that free-market ideology is a strong predictor of antivaccine sentiments; many libertarian parents oppose vaccinations, seeing them as infringing on parents' rights. (Despite popular perceptions, Lewandowsky found little evidence of a link between vaccine resistance and left-wing political views.) (SP)
> In uno studio pubblicato sulla rivista scientifica PlosOne, Lewandowsky osserva che negli Stati Uniti spesso c'è un legame tra l'ideologia liberista e il rifiuto dei vaccini. Molti sostenitori del libero mercato si oppongono ai vaccini perché li considerano un'imposizione dello stato che viola i loro diritti. Mentre, nonostante la percezione diffusa del contrario, Lewandowsky non ha trovato conferma di un legame tra la resistenza ai vaccini e le opinioni di sinistra.

Cultural transediting can further be achieved through omission or implicitation of parts of text that in Italian – a high context culture – would be regarded as superfluous. In the example below from *Project Syndicate*, "for their children's non-attendance" remains implicit in the target text as a different form of cohesion is created through possessive adjective reference – "i loro genitori" – instead of lexical repetition:

> Under a new law, unvaccinated children are not permitted to attend school, and parents of unvaccinated children can be fined for their children's non-attendance. (WPCK)
> In base alla nuova legge, i bambini non vaccinati non potranno frequentare la scuola, e per questo i loro genitori saranno passibili di multa.

In *Internazionale* translations, cultural transediting is particularly frequent in the "Vaccines on trial" article, where the translator needs to consider differences in the US and Italian legal systems and adapt the target text accordingly. This process is helped by the fact that its focus in Italian is on the evidence of the safety of vaccines, not on legal procedures in the US courts. Petitioner(s), for example, is rendered as 'ricorrente' or 'querelante' or glossed as in the excerpt below:

> Since its first case in 1988, the vaccine court has adjudicated more than 16,000 petitions and dismissed two-thirds of them. To the successful petitioners, and their lawyers, it has awarded about $3.6 billion. (VT)
> Da quando è entrato in funzione nel 1988, il tribunale dei vaccini ha giudicato più di 16mila richieste e ne ha respinte due terzi. Le persone che sono riuscite a ottenere l'indennizzo, e i loro avvocati, hanno incassato nel complesso 3,6 miliardi di dollari.

The *Project Syndicate* translations include two types of transcreation. One is the translation of the allusive headline "An advocacy of Dunces" with "Quando attivismo fa rima con egoismo", where the idiom *fare rima con* collocates with two words, *attivismo* and *egoism* that actually rhyme with each other and thus reinforce the idea. The other one is the decision not to translate the standfirsts of the articles into Italian – presumably an editorial choice that applies to all of them. By contrast, the *Science* article translations – as texts from an interspecialist journal translated for the lay public – are definitely the ones that present more instances of transcreation. Transcreation is found in the box "Da sapere. Come funzionano i vaccini" written by the Italian Health Research Institute (Istituto superiore di sanità) to provide readers with essential background information on vaccines from an authoritative source. Transcreation can also be seen at work in the title and standfirst of the articles. In *Internazionale* the overall purpose is to highlight that vaccines are necessary and safe by citing scientists as authoritative sources of information. The main article, "The Science of Persuasion", promotes information relating to the section in *Science* (News Features) to the title of the article, "La guerra dei vaccini", thus drawing on the war metaphor. The standfirst is much more extended as it provides an explicitation of the main points in the controversy over vaccines. The argument of interest to scientists, that is what they can do to persuade parents that vaccines are safe, is turned into a general reference to fighting misinformation and a statement on the authoritativeness of the journal the article comes from. In terms of cultural differences, this reflects the relevance of hierarchy in power relations in an Italian context:

Science	***Internazionale***
NEWS FEATURES Vaccine wars	**In copertina** **Kai Kupferschmidt, Science, Stati Uniti**
Vaccines save lives. But what is the most effective way to convince worried parents? *By Kai Kuperschmidt*	Proteggono dalle malattie e salvano milioni di vite, ma suscitano paura e diffidenza. Combattere la disinformazione è difficile: quali sono davvero i rischi dei vaccini? L'inchiesta di una delle più importanti riviste scientifiche del mondo
The science of persuasion	**La guerra dei vaccini**

Table 11
Transcreation of the title and standfirst of the main article from *Science* into Italian.

Similar strategies emerge in the translation of the *Science* article explaining the work of the US vaccine court. In Italian the bluntness of the title is mitigated by referring to trials indirectly through intertextuality as "Parola ai giudici" reminds readers of the standard phrase "la parola alla giuria" or of plays and films on trials such as "La parola ai giurati". Here, however, stress is laid again on authority and the last word is that of judges, not of lay people as in a jury. In Italian the standfirst is adapted in that the explicit English wording is replaced by a general description of what the US court does. This is another example of cultural mediation combining transcreation with cultural transediting, where the direct style typical of English is replaced with the required indirectness in Italian. In line with the overall purpose of these articles in Italian translation, a sentence is added to stress that in most cases, harm to people is not caused by vaccines, but by the way they are given the shots.

Science	***Internazionale***
Vaccines on trial	**Parola ai giudici**
The U.S. vaccine court weighs real versus bogus risks *By Meredith Wadman*	**Meredith Wadman, Science, Stati Uniti** Negli Stati Uniti esiste un tribunale che valuta le richieste di risarcimento per i danni delle vaccinazioni. Ma nella maggior parte dei casi si tratta di lesioni articolari provocate dall'ago.

Table 12
Transcreation of the title and standfirst of the "Vaccines on trial" article from *Science*.

In *Science* the third article is spread out over a number of pages in the form of boxes entitled "False" and followed by the myth in question. In *Internazionale* the boxes are turned into a single article, and a standfirst is added to stress once again that myths have no scientific ground. The titles of the individual boxes are used to introduce sections on the different myths. In this article, reference to the alleged harm caused by vaccines is again indirect, in line with the conventions of the target culture.

Science	***Internazionale***
Vaccine Myths	**I miti da sfatare** Negli anni si sono diffuse delle false credenze che non hanno fondamento scientifico. Per esempio la relazione dei vaccini con alcuni disturbi neurocomportamentali.
False: Vaccination can cause autism	Il legame con l'autismo
False: Mercury in vaccines acts as a neurotoxin	Danni da mercurio
False: Countering mercury from vaccines can make children better	La tesi di Mark Geier
False: Spreading out vaccines can be safer for kids	Distanziare le dosi
Lindzi Wessel	

Table 13
Transcreation of the false myth *Science* boxes in *Internazionale*.

As can be seen, each "False" headline is turned into a section title that can be easily understandable for the lay public. Consequently, the chemical process whereby mercury acts as a neurotoxin is simply translated as "Danni da mercurio" (mercury damage) whereas "Countering mercury…" is rendered with a reference to the doctor who promoted the theory dealt with in the article. Finally, the last "False" headline is shortened and much information is left implicit in Italian in a process that is adequate in the translation from a low-context to a high-context culture.

5. Conclusions

Public trust in immunization is quite high in English-speaking countries. By contrast, the attitude towards vaccine safety is particularly negative in Italy. Translation of articles on the controversy over vaccines needs to take account of culture-specific differences in sentiment. Analysis of corpora provided background information on how the debate over the controversial issue of vaccines and vaccine safety is framed in two culturally different contexts, the English-speaking countries and Italy. Information retrieved from comparison and contrast of modes of topic presentation in English and Italian was used as benchmark to identify translation strategies, assign them to the category of rewriting, transediting or transcreation and assess translation quality.

Instances of the main procedures for each category were detected and assigned a positive or negative score depending on the degree of success in rendering the feature(s) identified in the source text.

Results suggest that translations into Italian are mainly rewritten and transedited even when there is a major change in readership as in the *Science* articles published by *Internazionale*. Within the two categories, the most frequent procedures in both sets of translated articles are rearrangements – rewriting in the form of changes in phrase and clause order to achieve fluent prose – and situational transediting to suit the intended function(s) of the target text. Major instances of transcreation are only found in the *Science* articles translated by *Internazionale*. They are meant to recast the source texts to meet the requirements of a different readership.

Though the *Project Syndicate* articles in Italian translation feature the highest number of rewriting and transediting instances, many cases of adaptation are either unsuccessful or unnecessary. Albeit longer overall, *Science* articles in *Internazionale* translation show a comparatively sparing use of rewriting and transediting, which is altogether much more effective in realizing the intended function(s) of the target texts. In other words, *Science* articles in translation exhibit greater awareness of the advisability of recasting texts through rewriting, transediting and transcreation when a change in readership is involved – as regards both culture and expertise. In terms of quality assessment, *Internazionale* translations can be said to be better than *Project Syndicate* ones. However, further research is needed to establish why that is the case, especially as *Project Syndicate* does not provide information about its translation service other than the available languages.

Bionote: Maria Teresa Musacchio is Full professor of English language and translation at the University of Padova. Her research covers Translation Studies and specialised translation, Corpus-based/corpus-driven studies of specialised translation, special languages and terminology, Comparative linguistics of English and Italian special languages, Translation pedagogy, Translation-oriented terminology and terminography, and Issues in intercultural translation. Recently she has worked on the topic of communication and translation during emergencies from natural hazards, and was PI of the Padua unit of the EU-funded project Slándáil - Security System for Language and Image Analysis.
Virginia Zorzi holds a PhD in English Linguistics from the University of Padova. Her research interests are in Applied Linguistics and Public Communication of Science and Technology. Her research combines qualitative and quantitative approaches to the analysis of scientific and technological popularisation in newspapers.

Author's address: mt.musacchio@unipd.it; virginia.zorzi@unipd.it

References

Anthony L. 2017, *AntConc (Version 3.5.0)*, Waseda University, Tokyo. http://www.laurenceanthony.net/software (15.02.2019).

Baker M. 1996, *Corpus-based Translation Studies: The Challenges that Lie Ahead*, in Somers H.L. (ed.), *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan Sager*, Benjamins, Amsterdam/Philadelphia, pp. 175-186.

Baker P., Gabrielatos C., Khosravinik M., Krzyżanowski M., McEnery T. and Wodak R. 2008, *A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press*, in "Discourse & Society" 19 [3], pp. 273-306.

Baroni M., Bernardini S., Ferraresi A. and Zanchetta E. 2009, *The Wacky Wide Web: a Collection of Very Large Linguistically Processed Web-Crawled Corpora*, in "Language Resources and Evaluation" 43 [3], pp. 209-226.

Bianucci P. 2008, *Te lo dico con parole tue: la scienza di scrivere per farsi capire*, Zanichelli, Bologna.

Bondi M. 2010, *Perspectives on Keywords and Keyness*, in Bondi M. and Scott M. (eds.), *Keyness in Texts*, Benjamins, Amsterdam/Philadelphia, pp. 1-18.

Bonomi I. 2002, *L' italiano giornalistico. Dall'inizio del '900 ai quotidiani on line*, Firenze, Cesati.

Bucchi M. 2008, *Of Deficits, Deviations and Dialogues: Theories of Public Communication of Science*, in Bucchi M. and Trench B. (eds.), *Handbook of Public Communication of Science and Technology*, Routledge, London, pp. 57-76.

Byrne J. 2012, *Scientific and Technical Translation Explained*, St. Jerome Publishing, Manchester.

Calsamiglia H. and López Ferrero C. 2003, *Role and Position of Scientific Voices: Reported Speech in the Media*, in "Discourse Studies" 5 [2], pp. 147-173.

Calsamiglia H. and van Dijk T.A. 2004, *Popularization Discourse and Knowledge about the Genome*, in "Discourse and Society" 15 [4], pp. 369-389.

Chesterman A. 1997, *Memes of Translation*, Benjamins, Amsterdam/Philadelphia.

Chesterman A. 2004. *Hypotheses about Translation Universals*, in Hansen G., Malmkjaer K. and Gile D. (eds.), *Claims, Changes and Challenges in Translation Studies*, Benjamins, Amsterdam/Philadelphia, pp. 1-14.

De Mooij M. 2004, *Translating Advertising: Painting the Tip of an Iceberg*, in "The Translator" 10 [2], pp. 179-198.

Garzone G. 2006, *Perspectives on ESP and Popularization*, CUEM, Milano.

Gotti M. 2012, *La riscrittura del testo da specialistico a divulgativo*, in "Altre modernità" 11, pp. 145-159.

Halliday M.A.K. and Martin J.R. 1993, *Writing Science: Literacy and Discursive Power*, The Falmer Press, London/Washington D.C.

Hunston S. 2002, *Corpora in Applied Linguistics*, Cambridge University Press, Cambridge.

Hyland K. 2010, *Constructing Proximity: Relating to Readers in Popular and Professional Science*, in "Journal of English for Academic Purposes" 9 [2], pp. 116-127.

Katan D. 2003, *Translating Cultures*, St. Jerome Publishing, Manchester.

Katan D. 2013, *Intercultural Mediation*, in Gambier Y. and van Doorslaer L. (eds),

Handbook of Translation Studies, Benjamins, Amsterdam/Philadelphia, pp. 84-91.
Larson H.J., de Figueiredo A., Xiahong Z., Schulz W.S., Verger P., Johnston I.G., Cook A. R. and Jones N.S. 2016, *The State of Vaccine Confidence 2016: Global Insights through a 67-Country Survey*, in "EBioMedicine" 12, pp. 295-301.
Lefevere A. 1992, *Translation, Rewriting, and the Manipulation of Literary Fame*, Routledge, London/New York.
Molek-Kozakowska K. 2017, *Communicating Environmental Science beyond Academia: Stylistic Patterns of Newsworthiness in Popular Science Journalism*, in "Discourse and Communication" 11 [1], pp. 69-88.
Myers G. 1990, *Writing Biology: The Social Construction of Popular Science*, University of Wisconsin Press, Madison, WI.
Myers G. 1994, *Narratives of Science and Nature in Popularizing Molecular Genetics*, in Coulthard M. (ed.), *Advances in Written Text Analysis*, Routledge, London/New York pp. 179-190.
Myers G. 2003, *Discourse Studies of Scientific Popularization: Questioning the Boundaries*, in "Discourse Studies" 5 [2], pp. 265-279.
Olohan M. 2016, *Scientific and Technical Translation*, Routledge, London.
Olohan M. and Salama-Carr M. 2011, *Translating Science*, in "The Translator" 17 [2].
Pedersen D. 2014, *Exploring the concept of transcreation - transcreation as 'more than translation?'* in "Cultus" V [7], pp. 57-71.
Sager J.C. 1994, *Language Engineering and Translation. Consequences of Automation*, Benjamins, Amsterdam/Philadelphia.
Scott M. 1997, *PC Analysis of Key Words and Key Key Words*, in "System" 25 [2], pp. 233-245.
Scott M. and Tribble C. 2006, *Textual Patterns: Key Words and Corpus Analysis in Language Education*, Benjamins, Amsterdam/Philadelphia.
Scotto di Carlo G. 2014, *The Role of Proximity in Online Popularizations: The Case of TED Talks*, in "Discourse Studies" 16 [5], pp. 591-606.
Seguin È. 2001, *Narration and Legitimation: The Case of in Vitro Fertilization*, in "Discourse and Society" 12 [2], pp. 195-215.
Siegfried T. 2006, *Reporting from Science Journals*, in Blum D., Knudson M. and Henig R.M. (eds.), *A Field Guide for Science Writers*, Oxford University Press, Oxford, pp. 11-17.
Sinclair J. 1991, *Corpus, Concordance, Collocation*, Oxford University Press, Oxford.
Sinclair J. 2004, *Trust the Text: Language, Corpus and Discourse*, Routledge, London.
Stetting K. 1989, *Transediting – A new term for coping with the grey area between editing and translating*, in Caie G., Haastrup K., Jakobsen A.L. et al. (eds), *Proceedings from the Fourth Nordic Conference for English Studies*, University of Copenhagen, Copenhagen, pp. 371-382.
Stubbs M. 1996, *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*, Blackwell, Oxford.
Stubbs M. 2010, *Three Concepts of Keywords*, in Bondi M. and Scott M. (eds.), *Keyness in Texts*, Benjamins, Amsterdam/Philadelphia, pp. 21-42.
Ulrych M. 2015, *Traces of Mediation in Rewriting and Translation*, EduCatt, Milano.
Williams Camus J.T. 2016, *Get the metaphor right! Cancer treatment metaphors in the English and Spanish press*, in "Alfinge" 28, pp. 109-138.

PART III Transferring specialised knowledge to novices

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 511-533
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p511

A MULTIMODAL APPROACH TO TEACHING BUSINESS ENGLISH THROUGH FILMS
A case study[1]

VERONICA BONSIGNORI
UNIVERSITY OF PISA

Abstract – Audiovisual products, due to their intrinsic multisemiotic nature, in which each mode contributes to meaning making, have widely proven to be ideal material both for research purposes and for teaching. This is especially the case of films, which, despite their fictional character, have often been used in the context of language learning. As the concept of multiliteracies has become pivotal in language teaching, the use of films reflects a multimodal approach that can help students learn to exploit modes beyond verbal language (e.g., visual, gestural, spatial) to both understand and produce texts in the target language more effectively. This becomes particularly important in situated communicative contexts where domain-specific discursive, pragmatic and cultural features can create significant obstacles for language learners. The present paper focuses on a particular type of specialised discourse, i.e. business discourse, and intends to explore the ways in which films can be utilized as resource material in the English for Business and Economics class at the university. Some clips from a film representing this domain and with relevant communication exchanges that portray different business contexts and genres will be analysed both on the verbal and non-verbal level through a multimodal annotation software. This allows for the integration of verbal and non-verbal cues (e.g., gaze direction, hand/arm gestures, etc.) to show how a multimodal approach can actually help students to interpret a message, especially in ESP contexts.

Keywords: film; multimodality; ESP; business English; language teaching.

1. Introduction

Audiovisual products are intrinsically multisemiotic in nature, since each mode contributes to the meaning-making process (see, *inter alia*, Bateman, Schmidt 2012; Wildfeuer 2013). For this reason, they have widely proven to be ideal material both for research and teaching purposes. This is especially the case of films, which, despite their fictional character, have often been used in the context

[1] Research financed by the Italian Ministry for the University (PRIN 2015 no. 2015TJ8ZAS).

of language learning (see Sherman 2003; Kaiser, Shibahara 2014; Bruti 2015). Indeed, even though film language has been described as "written-to-be-spoken-as-if not-written" (Gregory, Carroll 1978), some recent studies have demonstrated its similarities to spontaneous conversation, especially in terms of authenticity (see Kozloff 2000; Forchini 2012; Bonsignori 2013). Moreover, films are effective tools in teaching contexts because they have been described as "an authentic source material (that is, created for native speakers and not learners of the language)" (Kaiser 2011, p. 233; see also Sherman 2003).

Teaching a foreign language using films entails exposing students to both aural and visual elements at the same time, thus giving them the chance to actually see how language is used in situated contexts, either general or specialised. In this way, they can be aware of the different varieties and accents used, they are able to listen to native speakers, and, most importantly, they can become aware of how meaning is made by the interplay between verbal and non-verbal cues such as arm/hand gestures, posture, gaze direction, head movements, facial expression, as well as intonation. Indeed, all these elements reflect the various semiotic resources that contribute significantly to supporting, integrating, and creating meaning in a given situational context. The meaning-making role of multiple semiotic modes beyond verbal language is a key concept introduced by the pioneering and seminal work of Kress and van Leeuwen (1996) and Lemke (1998), which has become the cornerstone of multimodal discourse analysis (see O'Halloran 2004; Scollon, Levine 2004; Norris 2004). More specifically, a communicative event cannot be investigated solely on the basis of the verbal element, but non-verbal elements should also be integrated into the analysis in a more holistic approach. As the concept of multiliteracies has become pivotal in language teaching (Kalantzsis, Cope 2013), the use of films reflects a multimodal approach (Jewitt 2014) that can help students learn to exploit other modes beyond that of verbal language (e.g., visual, gestural, spatial) in order to both understand and produce texts in the target language more effectively (O'Halloran *et al.* 2016). This becomes particularly important in communicative contexts where domain-specific discursive, pragmatic and cultural features can create significant problems for language learners.

Taking inspiration from a recent study which explores how to use clips from films and TV series representing different specialised domains to teach ESP (see Bonsignori 2018), the present paper focuses on a particular type of specialised discourse, i.e. business discourse, and intends to explore the ways in which films can be utilized as resource material in the English for Business and Economics class at the university. Interestingly, this type of audiovisual product has been widely employed to teach various topics relevant to the domain of business and management in general, starting with the work of Champoux (1999), who uses scenes taken from several films to teach organizational behaviour and management theories and concepts. Other examples are Briam

(2010), who shows how films can enhance the teaching of intercultural communication in business contexts; Pandey (2012), where films are used to teach cross-cultural management; and finally, O'Boyle and Sandonà (2014), in whose work they are used to teach business ethics.

This paper presents a case study based on a multimodal analysis of a selection of clips from the film *Margin Call* (2011, J.C. Chandor), which is included in a multimodal corpus of audiovisual texts for ESP settings that is currently under construction. The clips were chosen because they portray relevant communication exchanges in different business contexts and genres. These will be analysed both on the verbal and non-verbal levels, through a multimodal annotation software (see section 2). This makes it possible to integrate verbal and non-verbal cues to show how a multimodal approach can actually help students to interpret a message, particularly in specialised contexts in the business domain. The analysis is then followed by a section including some teaching applications and conclusions.

2. Methodology and data

The film chosen to represent the specialised discourse domain of Business and Economics is *Margin Call*[2] (2011, J.C. Chandor). It describes the day of a group of people working for MBS, a big and respected Wall Street investment bank, at the very beginning of the 2008 financial crisis. The fact that the story is primarily set in the offices of the bank and that the film director deliberately attempted to explain the subject in the simplest way possible through the characters' dialogues, without losing authenticity, made this film perfect for teaching purposes. The film was viewed in its entirety in order to select some scenes which were considered relevant for the teaching of English in a Business and Economics class. Seven clips were created with a maximum length of 4 minutes. These clips mostly show different business situations such as job interviews, meetings, conversations between colleagues, business negotiations, as well as the firing of employees. Apart from specialised vocabulary, in this case the analysis can also focus on the conversation strategies and the speech acts used, based on the situational context but also on register variation according to the participants' role, for instance, in an exchange between employer and employee *vs*. between colleagues (i.e., peer to peer). But, more importantly, the clips were analysed multimodally, that is, taking into account both verbal and non-verbal cues in order to verify how the interplay of different semiotic codes contributes to generating meaning and helps to interpret the message, especially in a specialised context such as business.

[2] For further information on this film: https://www.imdb.com/title/tt1615147/?ref_=nv_sr_1.

The multimodal analysis was carried out with the annotation software ELAN (Wittenburg *et al.* 2006),[3] with which it is possible to create an ad-hoc multi-tiered framework on the basis of the aims of the study. Table 1 below shows an example of the multi-levelled analytical structure for the present work.

TIERS	**CONTROLLED VOCABULARY**	
	Abbreviation	**Description**
Transcription		
Gesture_description	Ff	Forefinger forward
	Pu	Palm up
	OPdhT	Open palm down hitting the table
Gesture_function	indexical	to indicate position
	modal	to express certainty
	parsing	to mark different units within an utterance
	performative	to indicate the kind of speech act
	representational	to represent an object/idea
	social	to emphasize/highlight importance
Gaze	up	up
	back	back
	down	down
	left	left
	right	right
	out	out
Face	frowning	frowning
	smiling	smiling
Head	QN	quick nod
	MSsts	Moving slowly from one side to the other
Prosody	stress	paralinguistic stress
Notes	description of camera angles, audience reaction	

Table 1
Example of the multi-tiered analytical structure created in the ELAN software.

As can be noted, the left-hand column in Table 1 shows the various tiers, while the following column is devoted to the so-called "controlled vocabulary", namely the labels used to describe certain phenomena, which is divided into two sections, i.e., abbreviation and full description. Each clip was wholly transcribed so that the characters' speech appears in the Transcription tier. If the clip contained more than one interlocutor, a corresponding number of Transcription tiers were created in the software. This also enables the analyst to notice instances of interruptions, overlapping and so on. When gestures were used by interlocutors, these were annotated and described in the Gesture_description tier, using labels on the basis of Querol-Julián's (2011) model, whereas their functions were indicated in the Gesture_function tier, following Kendon (2004) and Weinberg *et al.*'s (2013) classifications. Other labels referring to information regarding gaze direction, facial expression and head movements were inserted in the corresponding tiers, as well as the

[3] ELAN was developed at the Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. It is freely available at: http://tla.mpi.nl/tools/tla-tools/elan/.

occurrence of prosodic stress on certain words in the Prosody tier. Finally, the Notes tier was used to annotate some information linked specifically to the film genre such as camera angles, since sometimes, when the character is speaking, the camera focuses on another character to show his/her reaction. Therefore, in these cases it is not possible to take into account the speaker's non-verbal cues in the analysis (see Bonsignori 2016).

For reasons of space, only two clips will be analysed in the following sections.

3. Analysis

3.1. Clip 1

The first clip is less than 4 minutes long (00:03:47) and shows a scene that takes place in Jared Cohen's conference room. Cohen is the Head of Fixed Income Security at MBS and has set up an emergency meeting with just a few other employees and the managers to discuss the sudden financial crisis that the company has suffered and to assess the entity of the damage. In the opening scene, we see Cohen with Sarah Robertson, who is the Chief Risk Management Officer, Ramesh Shah, a number cruncher, and David Horn, a firm lawyer, who are all waiting for Sam Rogers, an Executive, and his team, namely, Will Emerson, the Head of Trading, and two young analysts, one of whom is Peter Sullivan, the person that found out about the serious problems now facing the company. They arrive, enter the room and sit down around the table. It is 2 a.m.

The transcript of Clip 1 has been divided into three parts in order to facilitate its analysis. Generally, as can be noticed in the whole transcript, the situation is tense, so there are a few instances of sarcastic comments (turns 2, 10, 25, 33), and the language is quite informal (see the use of colloquialisms, e.g., *the math checks out* in turn 31, swear words, e.g., *fuck me* in turns 31 and 33, phrasal verbs and vague language, e.g., *put a few things in* in turn 14). Specialised vocabulary and expressions pertaining to the business domain are highlighted in italics, while the presence of gestures is indicated by the underlining of the linguistic items with which they co-occur. What follows is the first part of the transcript.

(1)

1	JARED COHEN to Sam	So, Sam, what do you have for us?
2	SAM ROGERS	It’ll be here in a minute. Finding somebody in the copy room at this hour was a little bit of a challenge.
3	JARED COHEN	Okay, let’s go right into the introductions. **This is Sarah Robertson, who you know, *Chief Risk Management Officer*, Ramesh Shah from upstairs. And David Horn, one of *the firm’s* in-house counsel.**
4	SAM	**Nice to meet you all. This is my *Head of Trading*, Will Emerson, and this is...**
5	WILL EMERSON	**Peter Sullivan and Seth Bregman. They work in our *Risk Department.***
6	JARED COHEN	Where’s Eric Dale?
7	SAM	He *was let go* today.
8	JARED COHEN (looking at Sarah)	Who do you have left in your Risk *Department*?
9	WILL EMERSON	As of today that would be Peter and our *Junior Analyst*, Seth.
10	JARED COHEN	Really?
11	COPY ROOM GUY entering (voice)	Will Emerson?
12	WILL EMERSON	Yes, please. Just hand them out.
13	JARED COHEN to Copy room guy	Thank you.
They all start reading the document		

The meeting opens with a series of introductions (turns 3-5), in bold, which are necessary in this context, since MBS is a big company, therefore people working in different departments often do not know each other. However, the managers and executives do know each other, as we can see with Sam Rogers and Jared Cohen, where Cohen addresses Rogers by his first name (turn 1), thereby expressing their familiar relationship. Since introductions represent an important speech act in the opening phase of meetings, Figure 1 shows the multimodal analysis of an introductory sequence, where Sam Rogers introduces his team, starting from Will Emerson (see turn 4 and the Transcription tier), using a gesture, labelled “Ff”, i.e., Forefinger forward, in the Gesture_description tier, with an indexical function, as indicated in the Gesture_function tier.

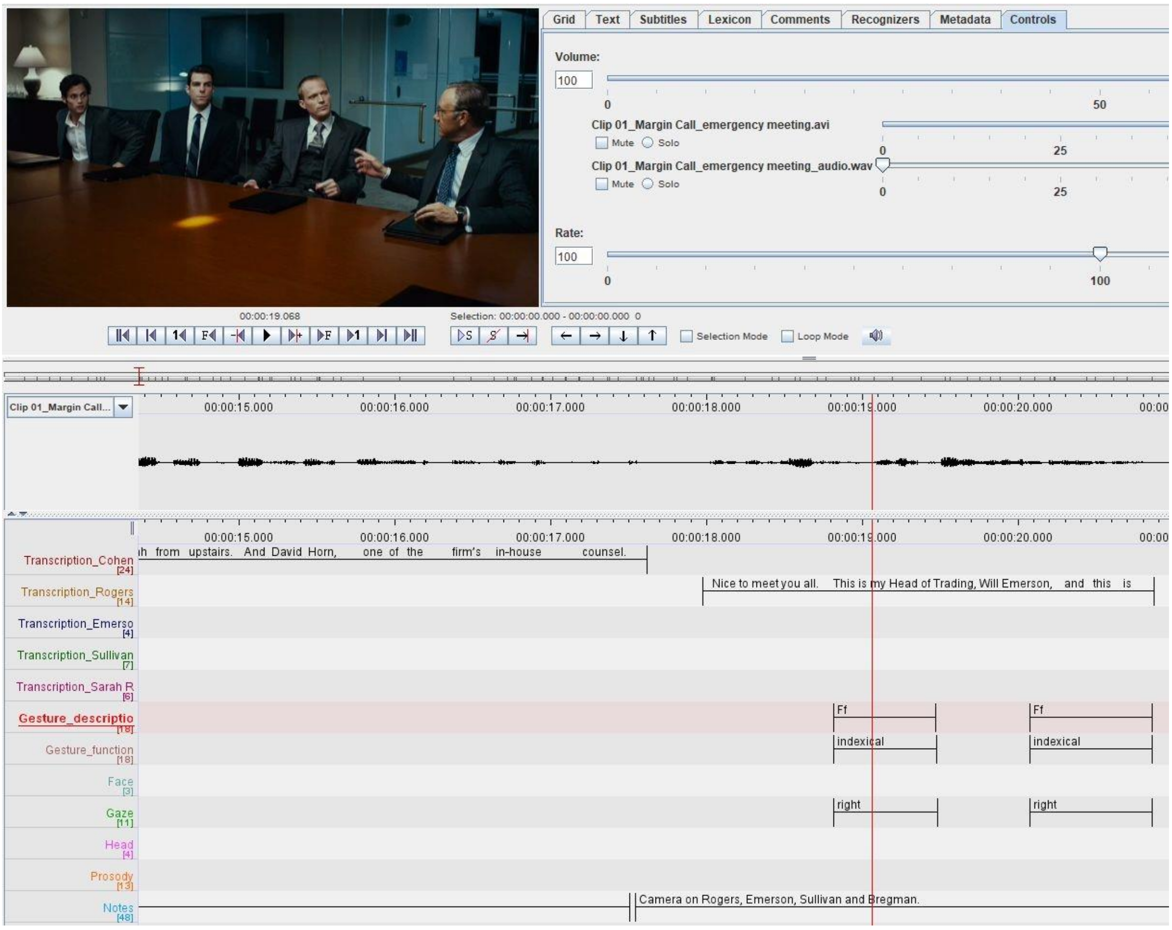


Figure 1
Gesture with an indexical function in an introduction sequence in a meeting (Clip 1).

The second part of the transcript starts with Sam Rogers trying to inform the participants about what has happened so far. Indeed, turn 14 is a good example of speaking about past events, with tense shifts (e.g., *had been working*, *left*), and using reported speech (e.g., *he asked him to take*). This report is then followed by a series of questions posed by Sarah Robertson to Peter Sullivan, the person who discovered the company's problems, in order to ascertain whether he is a reliable source. This is not a proper job interview, since Peter already has the job, but interestingly it looks like one. She asks him about his CV, his education, and then Cohen pops in to ask Peter why, despite his doctorate in engineering, he chose a job in the business sector (turns 19-28). These are the typical questions asked during a standard job interview.

(2)

14	SAM ROGERS	Apparently, Eric had been working on this for some time, but he wasn't able to finish it. This morning, as he left the building, he handed the program to Peter here and asked him to take a look at it. Peter did. He put a few things in that Eric seemed to be missing, and this is what came out.
15	SARAH to Peter	Peter, is this your work?
16	PETER	Mostly Mr. Dale's.
17	SARAH	But is this your draft?
18	PETER	Yes. Again, expanded on the original work by Mr. Dale. But, yes.
19	SARAH	**What's your background?**
20	PETER	**My background?**
21	SARAH	**Your CV?**
22	PETER	**I've been with the firm for two and a half years, working with Eric that whole time. But I hold a doctorate in engineering, specialty in propulsion, from MIT with a bachelor's from Penn.**
23	JARED COHEN to Peter	**What is a specialty in propulsion, exactly?**
24	PETER	**My thesis was a study in the ways that friction ratios affect steering outcomes in aeronautical use under reduced gravity loads.**
25	JARED COHEN	**So, you're a rocket scientist?**
26	PETER	**I was. Yeah.**
27	JARED COHEN	**Interesting. How did you end up here?**
28	PETER	**It's all just numbers, really. Just changing what you're adding up. And to speak freely, the money here is considerably more attractive.**
29	JARED COHEN	What time is it?
30	RAMESH SHAH	2:15.
31	JARED COHEN	Fuck me. Fuck me. And I'm guessing by the fact that you two haven't said anything that the math checks out.
32	SARAH	Look, we'd need some time to go over this. But Mr. Sullivan here seems like he knows what he's doing. So, it would appear we have a problem.
33	JARED COHEN	Oh, thank you for that. What time is it?
34	RAMESH SHAH	2:16.

In the third and final part of the transcript, Sam Rogers replies stiffly and confronts Jared Cohen when he suggests they cover everything up to solve the problem. Sam Rogers is deeply irritated by Cohen's implicit request and speaks using a sarcastic tone. In turn 40, he says something that his interlocutor knows perfectly well, being a professional like him. Nevertheless, he feels the urge to explain the situation very clearly, by using prosodic stress, gestures and a modified idiom (see he uses the verb *hide* instead of *sweep* with *something under the rug*) in order to get a clear picture of the situation.

(3)

35	JARED COHEN (stands up) to Sam	Fuck me. Fuck me. Sam, *how long under normal operations* <u>*would it take*</u> *your people* <u>*to clear that*</u> *from our books?*
36	SAM	What? All of it?
37	JARED COHEN	Yes.
38	SAM	I don't know. Weeks.
39	JARED COHEN	Weeks?
40	SAM	Yeah. Weeks. But you certainly know that *our business is* <u>*selling and buying*</u>? It doesn't work for very long without both components. We suddenly stop buying for a day or two**, that's not something you can hide under the rug**. **That gets out. And when it does, <u>this whole thing</u> <u>comes to an end</u>, and <u>right quick</u>.**
41	JARED COHEN	I understand.
42	SAM	Do you?
43	JARED COHEN	Yes. How many *traders* do we have left between your floor and Peterson's?
44	SAM To Will	I don't know, what's...
45	WILL	60.
46	RAMESH SHAH	Jared, as <u>I look at this</u> <u>more closely</u>, <u>it is these *VAR numbers*</u> that are setting this thing off--.
47	JARED COHEN to all (leaves)	Excuse me.

More specifically, Figures 2 and 3 show the multimodal analysis of the explanatory sequence in turn 40. In Figure 2, Sam Rogers' sarcastic utterance *But you certainly know that our business is selling and buying* is accompanied by the gesture visible in the still image and labelled in the Gesture_description tier "PsdmStS", i.e., Palms down moving from side to side, thus performing a representational function. Moreover, the key words *selling and buying* all bear prosodic stress, see Prosody tier, to highlight that both components are needed.

Finally, in Figure 3, the Gesture_description tier shows that three different types of gestures are used to iconically represent the metaphor Sam Rogers uses to clarify his point. The still image portrays the last gesture labelled "Psmd", i.e., Palms moving down. All three gestures perform a representational function and, again, key words are stressed.

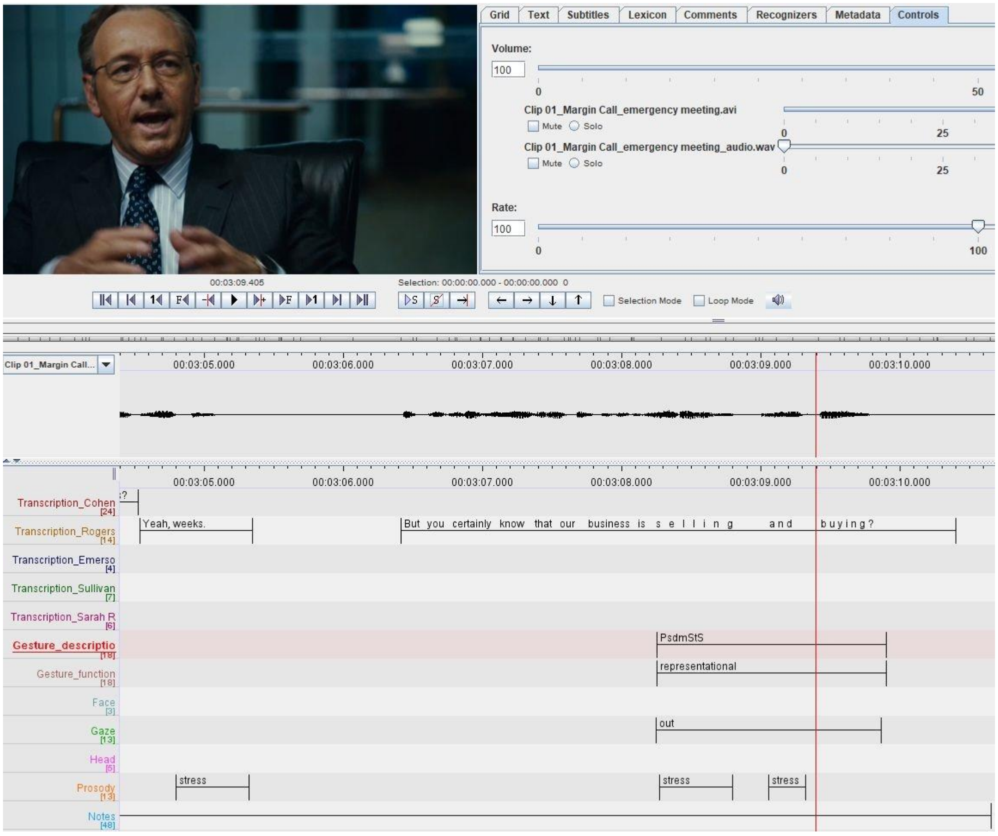


Figure 2
Gesture with a representational function
in an explanatory sequence in a meeting (Clip 1).

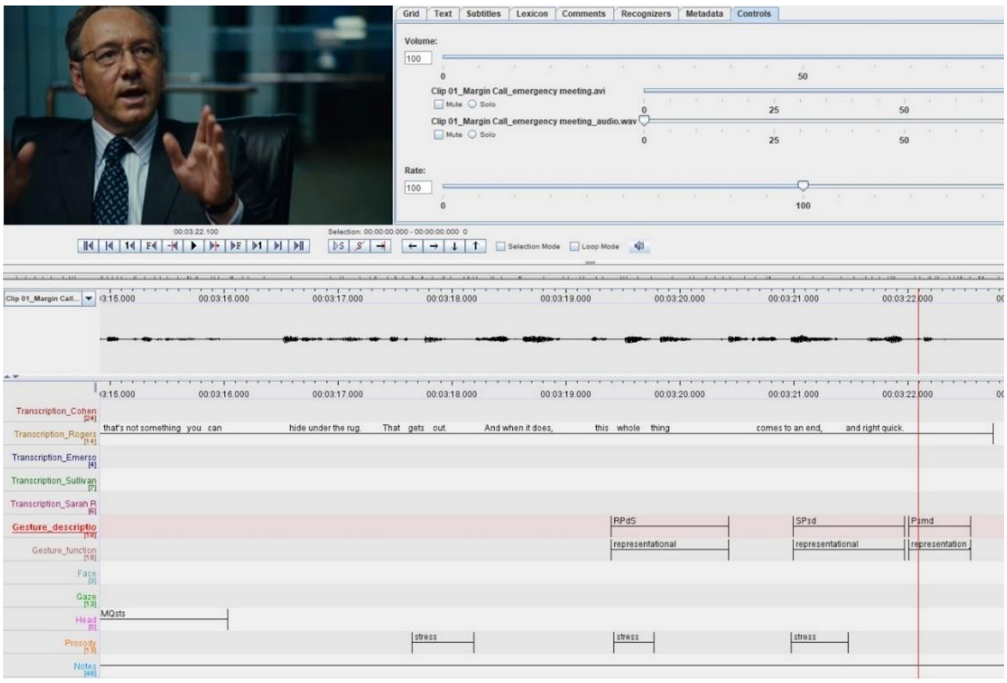


Figure 3
Gesture with a representational function accompanying a metaphor (Clip 1).

Table 2 below summarises the gestures used by the various participants in the meeting and their main functions. As can be noted, out of a total of 18 gestures, six, that is, one third, accompanied the introduction sequence. Moreover, gestures mainly performed indexical and representational functions, thus suggesting that the main communicative function expressed in this context was one of explanation and clarification, with gestures helping visualise important concepts and identify a specific person (see as in introductions).

N	Transcript	Gesture - abbrev	Gesture - detailed description	Gesture - function
1	This is Sarah Robertson, who you know….	Ff	Forefinger forward	indexical
2	Chief Risk management Officer.	Ff	Forefinger forward	indexical
3	This is my Head of trading…	Ff	Forefinger forward	indexical
4	And this is…	Ff	Forefinger forward	indexical
5	Peter Sullivan, …	Pu	Palm up	indexical
6	Peter and our junior analyst, Seth	Pu	Palm up	indexical
7	Just hand them out.	FP	forefinger pointing out	performative
8	[PAUSE] and this is what came out!	OPdhT	Open palm down hitting the table	social
9	Again, … expanded on the original work by Mr. Dale.	FTemC	finger and thumb extended moving in circle	representational
10	I've been with the firm…	OPu	opening palm up	performative
11	Specialty in propulsion…	OPu	opening palm up	performative
12	We'd need some time to go over this	PDmc	palm down moving in circle	representational
13	But Mr. Sullivan here seems to know what's he's doing.	Pupi	palm up pointing the interlocutor	indexical
14	… would it take your people to **clear that** from our books?	FTembf	forefinger and thumb extended moving back and forth	representational
15	Our business is **selling and** buying?	PsdmStS	palms down moving from side to side	representational
16	This **whole** thing	RPdS	raising palm down and separating	representational
17	Comes to an **end**	SPsd	separating palms down	representational
18	And right quick	Psmd	palms moving down	representational

Table 2
Summary of gestures in Clip 1.

3.2. Clip 2

The second clip selected for this study is the same length as Clip 1, and shows a meeting with board members. More specifically, due to the situation of emergency brought up by the previous restricted meeting shown in Clip 1, a new emergency meeting has just been set up and a lot of important people are waiting for their boss, John Tuld, in the CEO's executive boardroom. The clip starts with John Tuld entering the room, greeting a few people, and then sitting down at the head of the table and starting to ask questions in order to understand the huge problems his company is in. Although there are several people attending the meeting, the dialogic exchange is between John Tuld, Jared Cohen, Head of Fixed Income Securities, and Peter Sullivan, analyst and associate. What follows is the first part of the transcript of Clip 2, which has been divided into two parts. The occurrence of gestures is indicated by underlining the verbal items with which they co-occur.

(4)

1	JOHN TULD to all	[Please, sit down. Welcome, everyone. I must apologize for dragging you all here at such an uncommon hour. But from what I've been told, this matter needs to be dealt with urgently.] So urgently, in fact, it probably should have been addressed weeks ago. But that is spilt milk under the bridge. So, why doesn't somebody tell me what they think is going on here?
2	JARED COHEN	Mr. Tuld, **as I mentioned earlier**, if you compare the figure at the top of page 13...
3	JOHN TULD	Jared, it's a little early for all that. Just speak to me in plain English.
4	JARED COHEN	Okay.
5	JOHN TULD	In fact, I'd like to speak to the guy who put this together. Mr. Sullivan, is it? Does he speak English?
6	JARED COHEN	Sir?
7	JOHN TULD	I'd like to speak with the analyst who seems to have stumbled across this mess.
8	JARED COHEN	**Certainly**. That would be Peter Sullivan. Right here.
9	JOHN TULD	Oh, Mr. Sullivan, you're here! Good morning! *Maybe you could tell me what you think is going on here.* And please, speak as you might to a young child or a golden retriever. It wasn't brains that got me here. I can assure you of that.

This sequence is exemplary as, in the first place, it shows some key features regarding the structure of a meeting – i.e., an aspect that is often taught in the Business and Economics class – namely, how to open a meeting, with a greeting, followed by an apology and the reason/topic of the emergency meeting itself (turn 1); the role of chair, here played by the boss, John Tuld, who has to manage the discussion, avoiding overlapping and interruptions; and some strategies used to pass the floor (turn 8), here accompanied by a gesture performing an indexical function, as shown in Figure 4.

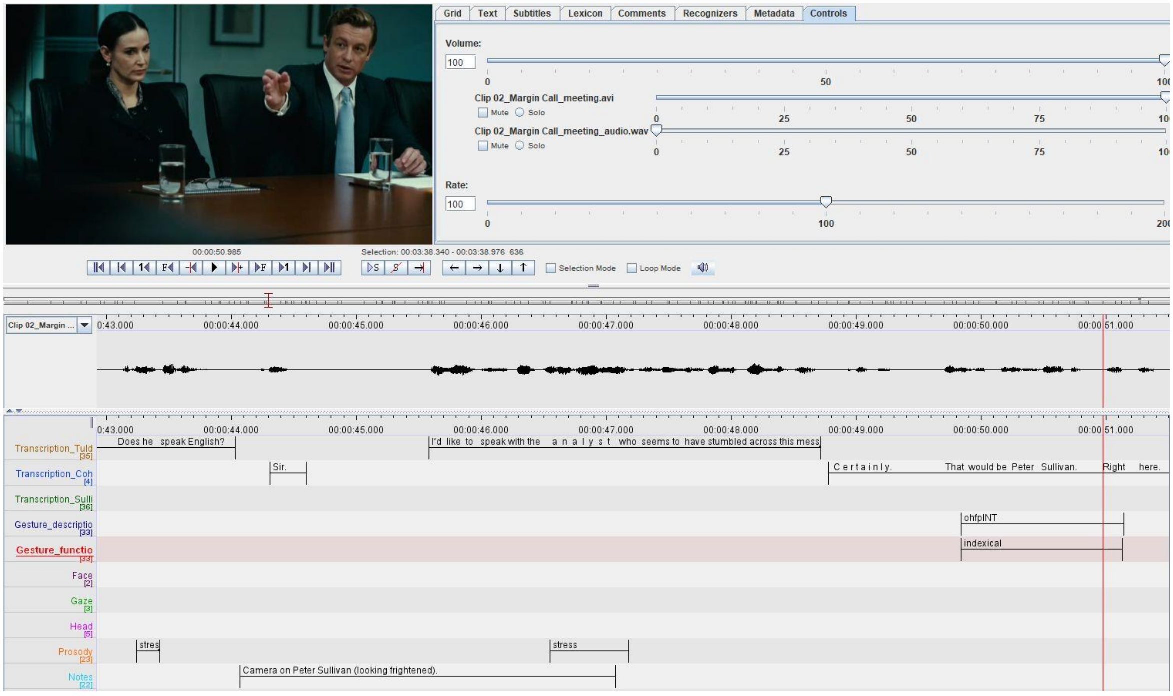


Figure 4
Gesture performing an indexical function to pass the floor in a meeting (Clip 2).

The second part of the clip (and transcript) is explanatory, as Peter Sullivan is asked to explain what the problem is.

10	PETER	Well... Oh uh... sir, **as you may or may not know**, I work here for Mr. Rogers as an associate in the Risk Assessment and Management Office at MBS.
11	JOHN TULD	Please. Just relax. Stand up. Tell us in a clear voice. *What is the nature of the problem?*
12	PETER	Okay. Uh... Well, **as you probably know**, over the last 36 to 40 months the firm has begun packaging new MBS products that combine several different tranches of rating classifications in one tradable security. This has been enormously profitable, **as I imagine you noticed**.
13	JOHN TULD	I have.
14	PETER	**Well**, the firm is currently doing a considerable amount of this business every day. Now the problem, which is, **I guess**, why we are here tonight, is that it takes us, the firm, about a month to layer these products correctly, thereby posing a challenge from a Risk Management standpoint.
15	JOHN TULD	And, Mr. Sullivan, *that challenge is?*
16	PETER	**Well**, we have to hold these assets on our books longer than we might ideally like to.
17	JOHN TULD	Yes.
18	PETER	But the key factor here is these are essentially just mortgages. So that has allowed us to push the leverage considerably beyond what you might be willing or allowed to do in any other circumstance, thereby pushing the risk profile without raising any red flags.
19	JOHN TULD	Now... Thank you, Mr. Sullivan. Sit down. What I'm guessing your report here says, and give me some rope here, what I'm guessing it says is that considering the, shall we say, bumpy road we've been on the last week or so, that the figures your brilliant co-workers up the line ahead of you have come up with don't make much sense anymore, considering what's taking place today.

20	PETER	Actually, not what's taking place today, but what's already taken place over the last two weeks.
21	JOHN TULD	So, you're saying this has already happened?
22	PETER	**Sort of.**
23	JOHN TULD	Sort of. And, Mr. Sullivan, what does your model say that that means for us here.
24	PETER	**Well**, that's where it becomes a projection. But... uhm…
25	JOHN TULD	You're speaking with me, Mr. Sullivan.
26	PETER	**Well**, sir, if those assets decrease by just 25% and remain on our books, that loss would be greater than the current market capitalization of this entire company.

Generally, the discussion is articulated in a sequence of questions, mostly asked by John Tuld, and responses. Various types of questions can be found in this extract (highlighted in italics), ranging from indirect (turn 9) and direct questions (turn 11) to an assertion turned into a question through a final rising intonation (turn 15). Moreover, the variations in register are also significant, especially in the way they reflect the characters' roles. More specifically, John Tuld, the boss, tends to use quite an informal register, e.g., the use of the mixed idiom in turn 1, *spilt milk under the bridge*, a blend between 'it's no use crying over spilt milk' and 'water under the bridge', and of the idiom in turn 19 *give me some rope here*; colloquial terms and expressions such as *drag* (turn 1) or *stumble across this mess* (turn 7); metaphors such as *bumpy road we've been on* (turn 19), which in some way clashes with his important role in the company. However, at the same time, his casual style highlights his aggressiveness,[4] which is also expressed through ironic and sarcastic comments (e.g., *Does he speak English?* in turn 5), and a request for clarity and simplicity, also through self-deprecation (e.g., *And please, speak as you might to a young child or a golden retriever. It wasn't brains that got me here* in turn 9). He also addresses his subordinate Cohen by his first name, *Jared* (turn 3), while Cohen calls him *Mr Tuld* (turn 2) or *Sir* (turn 6), thus acknowledging his higher status. In the same way, both Cohen and Sullivan, who both are Tuld's employees, even though at different levels, use mitigation strategies in the form of hedges (highlighted in bold) when they are asked to give explanations by their boss, in the attempt to not sound arrogant, e.g., *as you may or may not know* (turn 10). Sullivan performs the difficult task of explaining the problems the firm is having by using simple but specialised vocabulary, and several pauses and gestures. Indeed, one of his crucial turns in the transcript (turn 18) is almost entirely accompanied by the use of gestures, which happen to be an extremely valuable tool in the meaning-making process. An example is the use of the gesture labelled as "oHm1s", standing for 'open hand moving to one side', which is performed while uttering the word *push*,

[4] John Tuld also speaks with a British accent, which contributes to strengthening his mean demeanor, as happens with villains in animated films. Indeed, giving a British accent to villains is a convention in American cinema, which has lead to stereotypically associate it with meanness.

thus iconically representing the meaning of the verb itself and helping not only Mr Tuld but also viewers (and students) to understand the verbal message (Figure 5).

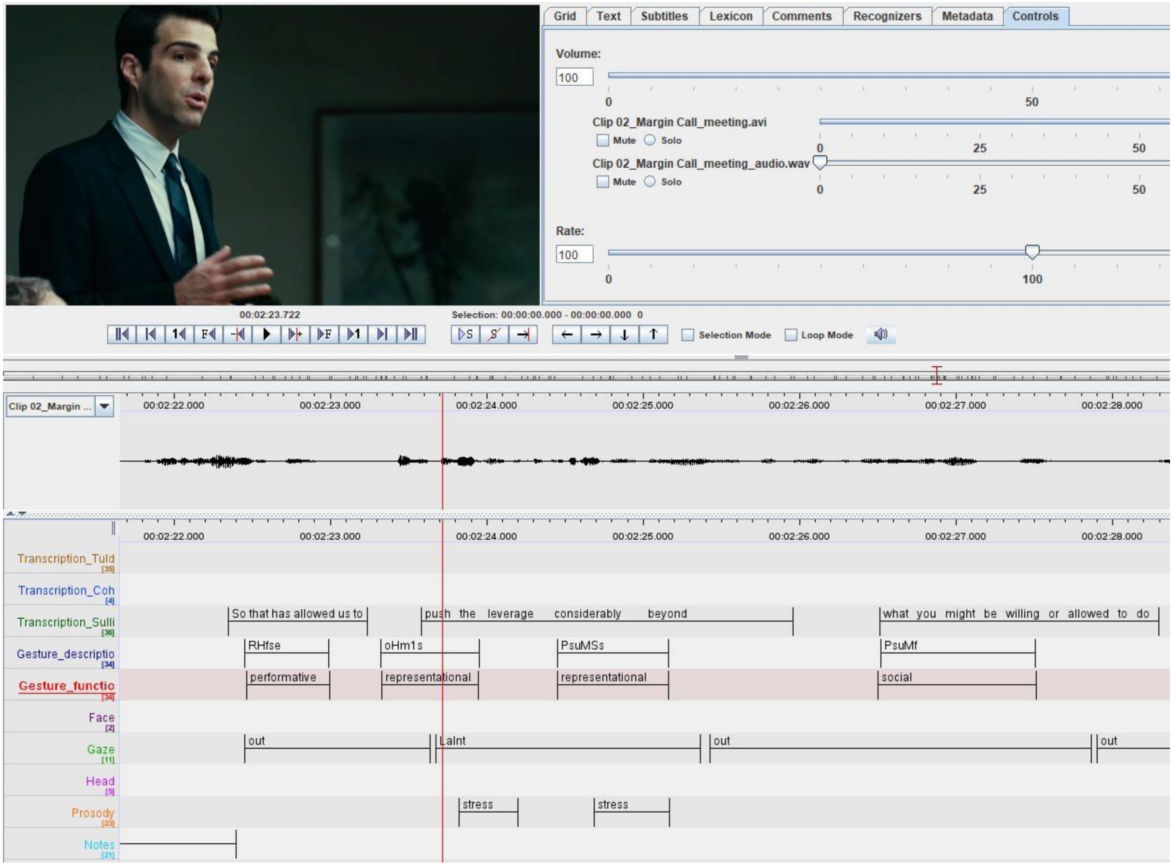


Figure 5
Gesture performing a representational function in an explanatory sequence (Clip 2).

Table 3 lists the gestures used by the speakers in this meeting and the main functions they perform. It is worth pointing out that the number of gestures in this clip is nearly double those in Clip 1, despite the same clip duration – i.e., 34 in Clip 2 vs. 18 in Clip 1.

N	Transcript	Gesture – abbrev.	Gesture – detailed description	Gesture – function
1	Jared, it's a little early for all that. Just speak to me in plain English.	PdDC	palm down describing circles	performative
2	I'd like to speak with the guy who put this together.	LOBJu	lifting object up	indexical
3	That would be Peter Sullivan. Right here.	oHfpINT	open hand forward pointing interlocutor	indexical
4	--	RoH	raising open hand	performative
5	And **please**,	PdMf	palm down moving forward	parsing
6	speak	PdMf	palm down moving forward	parsing
7	as you **might**	PdMf	palm down moving forward	parsing
8	to a **young** child	PdMf	palm down moving forward	parsing

9	Sir, as you may	PuMf	palm up moving forward	indexical
10	or may not know	PuMS	palm up moving to one side	indexical
11	I work here for Mr. Rogers	PdMS	palm down moving to one side	indexical
12	At MBS.	OPsu	opening palms up	performative
13	Just relax. Stand up.	FPsMu	facing palms moving up	performative
14	Tell us in a clear voice.	PsdPMSs	palms down parting and moving to the sides	performative
15	Well, as you probably know,	PdMf	palm down moving forward	indexical
16	over the last thirty-six to forty months	PsuMSs	palms up moving to the sides	social
17	**combine**	PsuMSs	palms up moving to the sides	social
18	in **one** tradable security.	FPsd	facing palms down	representationa l
19	as I imagine	PuMf	palm up moving forward	indexical
20	Well, the firm is currently doing a **considerable** amount of this…	PsuMSs	palms up moving to the sides	social
21	Now the **problem**,	PuMS	palm up moving to one side	social
22	Well, we have to hold…	PuMS	palm up moving to one side	modal
23	So that has allowed us to	RHfse	rotating hand thumb & forefinger extended	performative
24	**push** the leverage	oHm1s	open hand moving to one side	representationa l
25	**considerably**	PsuMSs	palms up moving to the sides	representationa l
26	what you might be willing or allowed to do	PsuMf	palms up moving forward	social
27	in **any** other circumstance,	PdMS	palm down moving to one side	representationa l
28	**thereby**	PsuMSs	palms up moving to the sides	social
29	Now, thank you, Mr. Sullivan.	FPd	forefinger pointing down	performative
30	and give me some rope here	oHFsMqbf	open hand with fingers moving quickly back and forth	performative
31	What I'm guessing it says	FTE	forefinger touching eyebrow	modal
32	What does your model say that that means	SN	scratching nose	social
33	You're speaking with **me**, Mr. Sullivan.	oHsMb	open hands moving back	indexical
34	That loss… would be greater than the current market capital (…)	oHm1s	open hand moving to one side	social

Table 3
Summary of gestures in Clip 2.

Most gestures are used by two characters, namely Peter Sullivan and the CEO, John Tuld. Out of a total of 34 gestures, 19 are used by Sullivan in his explanatory sequences, while 12 are used by Tuld accompanying directives, in the form of orders and requests. Therefore, both verbal and non-verbal cues contribute in representing the role of these two characters.

4. Teaching applications

The two clips selected from *Margin Call* both show the context of a company meeting and offer different suggestions for their uses in the a Business and Economics class. Taking Clip 2 as an example, it is always worth pointing out the importance of setting the background. First, a brief synopsis of the film should be provided, since not all students may be familiar with the story. Second, the scene shown in the clip should be introduced as well by describing the situational context, who the participants are, their role, etc. Then, students should just watch the clip once in order to get an idea of the main content, and watch it a second time to do a listening/watching comprehension exercise, e.g., a completion exercise where the missing items could be specialised vocabulary, parts of idiomatic expressions and some verb forms. Correction could be made either by watching the clip again with pauses and with the subtitles in English, or the instructor could simply read through the complete transcript, stressing the words that were originally omitted. Finally, students could do a comprehension exercise with True/False questions to be corrected later on.

At this point, attention should be paid to the meeting context and to the participants' roles by analysing not only verbal but also, and especially, non-verbal cues. Various types of activities can be designed in order to raise learners' awareness of the import of non-verbal communication. An example could be to divide the class into different groups and assign each one a non-verbal cue such as gestures, head movements, gaze direction, facial expression, etc., on which to focus while viewing the clip. Then, each group should watch the clip, with the instructor pausing it at relevant moments in order to give students enough time to note down the description of the assigned visual element in their own words. For this purpose, they could utilize a prepared form as illustrated in Table 4 below for gestures.

N	Image	Transcript	Gesture description + interpretation
1		Well, we have to hold…	(*) *Palm up moving to one side → modal function: showing self-confidence*
2		So that has allowed us to	
3		push the leverage	
4		considerably	
5		what you might be willing or allowed to do	
6		in any other circumstance,	
7		thereby…	

Table 4
Example of activity based on the analysis of gestures
in an explanatory sequence of Clip 2.

Due to the high occurrence of gestures and other non-verbal elements, Clip 2 should be divided into separate scenes. Thus, the form in Table 4 refers to one of the explanatory sequences in which Sullivan attempts to explain the problems of the company to his boss, Tuld. As can be noted from the third column, students are asked not only to describe the gestures portrayed in the still images shown in the left-hand column, but they should also think about the functions they perform and their effect on the interlocutor. In order to make the task easier, the corresponding transcript is also made available in the central column. An example of a possible student answer is provided for the first frame. The student might first describe the gesture physically (i.e., "palm up moving to one side"), and then attempt to interpret its function in the context. In this case, it appears to have a modal function to express the speaker's self-confidence while suggesting what could be done to save the company. Sullivan, who is intimidated by his boss, however, has to show him his expertise and that he perfectly knows what he is talking about. This gesture, together with his direct gaze towards his interlocutor, contributes to achieving this effect.

Ideas should be then discussed as a class, and results could also be shown by the instructor using the ELAN software, thus integrating the analysis of the different semiotic resources altogether. Figure 6 shows the multimodal analysis of the explanatory sequence in question, with annotations regarding gestures, gaze direction, and prosodic stress.

With the possibility of actually watching and listening to the clip while showing the multi-tiered analysis, ELAN allows for a thorough investigation of communicative events. In this way, students have all the tools needed to be able to consider how non-verbal communication contributes to shaping the characters' stance and its role in the meaning-making process.

After this in-depth multimodal analysis, the comprehension exercise with True/False questions should be completed again and, this time, corrected in order to verify to what extent the multimodal analysis of the whole clip helped the students to understand it. Moreover, other linguistic phenomena that have emerged in Clip 2 can be analysed as well, i.e., idioms, speech acts, etc., based on the topics of the course, and particular attention should be paid to specialised vocabulary, perhaps with an exercise to trigger meaning.

Finally, the clip itself can also be used for other types of language learning activities. For instance, on the lexical level, since the clip is about a company meeting, collocations with the lexeme 'meeting' could be provided, as well as a list of recurring phrases and expressions used to open, run and close a meeting, with special focus on the role of the chairperson. On the grammar level, exercises on one of the wide range of grammar topics that can be found in the transcript could be done: the passive voice, question types, modal verbs, tense revision, etc.

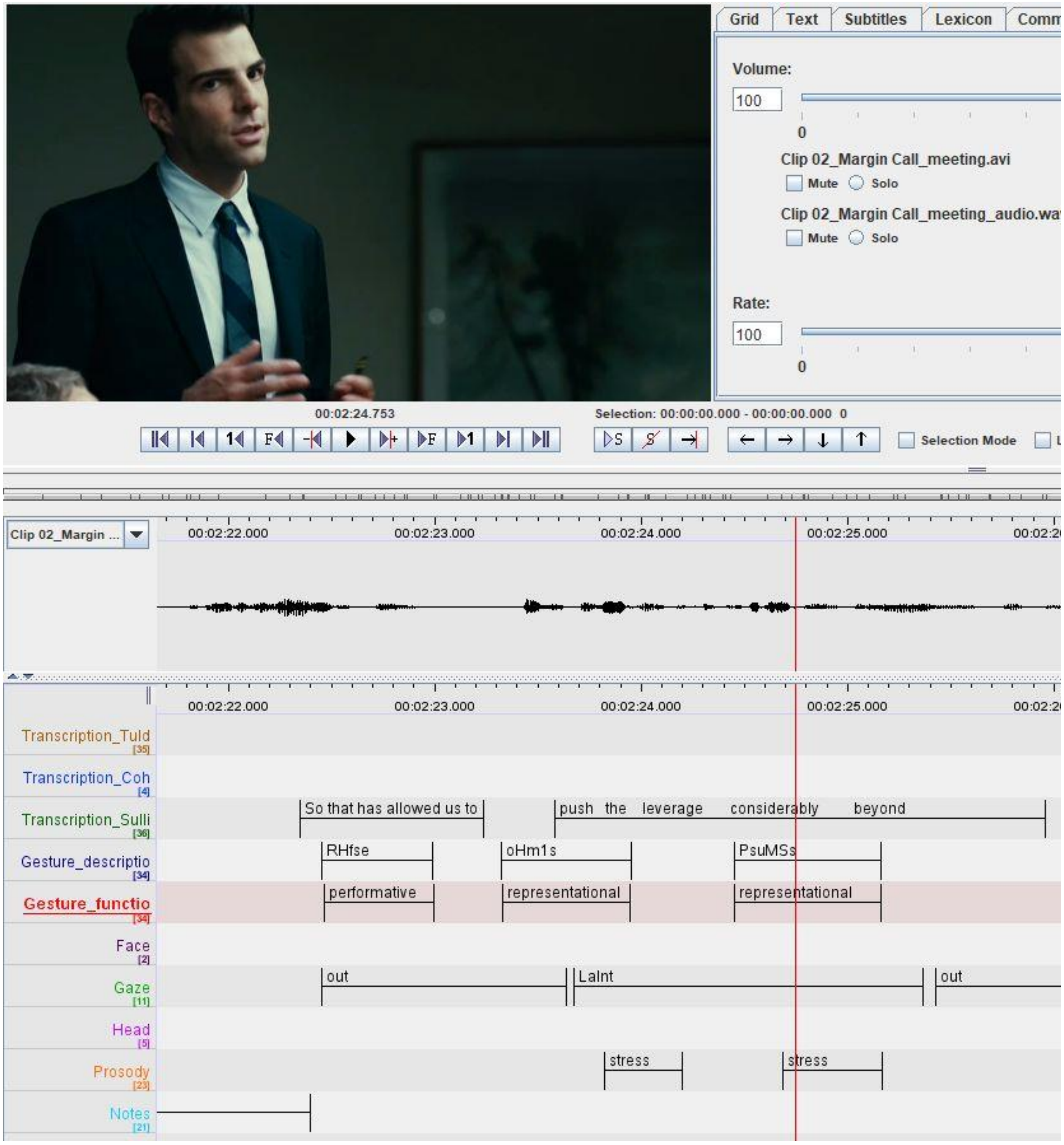


Figure 6
Example of the multimodal analysis of a scene of Clip 2.

This is just an example of a lesson plan based solely on Clip 2, but there are also other options. One could use both Clip 1 and 2 in combination, since they provide a representation of two different types of meetings. In this case, the focus of the lesson could be on the participants' role in each meeting with special attention to register shift from the restricted meeting in Clip 1 to the board meeting in Clip 2 – i.e., the casual style of interlocutors in Clip 1, where the informal context allows them to be even blunt, as in turn 28, where Peter Sullivan admits that his choice to work in a bank is solely motivated by salary *vs*. the employees' more controlled way of speaking to the boss in Clip 2.

5. Conclusions

The present work has shown the richness of the film genre as a multimodal product that can be used as a teaching resource also in specialised contexts such as a Business and Economics class. In particular, it is better to use film clips rather than the entire film (see, *inter alia*, Moskovich, Shaf 2012; Donnelly 2014; Baños, Bosch 2015), because this would be far too demanding for students. Conversely, short clips allow for in-depth analysis and give students enough time to concentrate on specific aspects or phenomena. The advantages of using film clips in the classroom are multiple. First, they provide models of spoken language, in general, and of specialised discourse in context, in particular, i.e., the business meeting in this case study. For this reason and for their plausibility as natural communicative events, they also "have the potential to engage students' interest" (Timmis 2005, p. 118). Second, they represent a valid chance for students to be exposed to the language spoken by native speakers, thus becoming aware of different accents and varieties, e.g. British English spoken by John Tuld *vs*. American English spoken by all the other characters in Clip 2. Third, being multimodal, film clips are the perfect tool to raise students' awareness of the importance of non-verbal cues such as gestures, posture, gaze direction, as well as intonation, in the meaning-making process. Finally, they allow for more practical activities and exercises based on the selected clips, for example to trigger grammar rules, specialised vocabulary, and so on. Last but not least, they can also be used to introduce key concepts according to the topics of the course, e.g., the structure of the meeting and its main communicative functions in the Business class.

Bionote: Veronica Bonsignori is a Researcher in English Language and Linguistics at the University of Pisa, Department of Philology, Literature and Linguistics. She holds a PhD in English Linguistics (2007) from the University of Pisa, where she teaches English for Tourism and Business English. Her interests are in the fields of pragmatics, sociolinguistics, audiovisual translation, multimodality, and ESP. She has contributed to national and international conferences and has published several articles, focusing on the transposition of linguistic varieties in Italian dubbing (2012, Peter Lang), and on linguistic phenomena of orality in English filmic speech in comparison to Italian dubbing, as in the paper co-authored with Silvia Bruti on introductions and wishes (2016, Equinox). She also authored the monograph *English Tags: A Close-up on Film Language, Dubbing and Conversation* (2013, Cambridge Scholars Publishing) and co-edited with Belinda Crawford Camiciottoli the volume *Multimodality Across Communicative Settings, Discourse Domains and Genres* (2016, Cambridge Scholars Publishing).

Author's address: veronica.bonsignori@unipi.it

References

Baños J.E. and Bosch F. 2015, *Using Feature Films as a Teaching Tool in Medical Schools*, in "Educación Médica" 16 [4], pp. 205-242.

Bateman J.A. and Schmidt K.H. 2012, *Multimodal Film Analysis: How Films Mean*, Routledge, London.

Briam C. 2010, Outsourced: *Using a comedy film to teach intercultural communication*, in "Business and Professional Communication Quarterly" 73 [4], pp. 383-398.

Bonsignori V. 2013, *English Tags: A Close-up on Film Language, Dubbing and Conversation*, Cambridge Scholars, Newcastle upon Tyne.

Bonsignori V. 2016, *Analysing Political Discourse in Film Language: A Multimodal Approach*, in Bonsignori V. and Crawford Camiciottoli B. (eds.), *Multimodality across Communicative Settings, Discourse Domains and Genres*, Cambridge Scholars, Newcastle upon Tyne, pp. 189-211.

Bonsignori V. 2018, *Using Films and TV Series for ESP Teaching: A Multimodal Perspective*, in Crawford Camiciottoli B. and Campoy Cubillo M.C. (eds.) *Multimodal Perspectives on English Language Teaching in Higher Education*, Special Issue of "System", 7 (2018), pp. 58-69.

Bruti S. 2015, *Teaching Learners How to Use Pragmatic Routines through Audiovisual Material*, in Crawford Camiciottoli B. and Fortanet-Gómez I. (eds.), *Multimodal Analysis in Academic Settings*. *From Research to Teaching*, Routledge, London, pp. 213-236.

Champoux J.E. 1999, *Film as a Teaching Resource*, in "Journal of Management Inquiry" 8 [2], pp. 206-217.

Donnelly D. 2014, *Using Feature Films in Teaching Historical Understanding: Research and Practice*, in "Agora" 49 [1], pp. 4-12.

Forchini P. 2012, *Movie Language Revisited. Evidence from Multidimensional Analysis and Corpora*, Peter Lang, Bern.

Gregory M. and Carroll S. 1978, *Language and Situation: Language Varieties and Their Social Contexts*, Routledge & Kegan Paul, London.

Jewitt C. (ed.) 2014, *The Routledge Handbook of Multimodal Analysis*, Routledge, London.

Kaiser M. 2011, *New Approaches to Exploiting Film in the Foreign Language Classroom*, in "L2 Journal" 3 [2], pp. 232-249. http://escholarship.org/uc/item/6568p4f4 (22.10.2018).

Kaiser M. and Shibahara C. 2014, *Film as a source material in advanced foreign language classes*, in "L2 Journal" 6 [1], pp. 1-13. http://escholarship.org/uc/item/3qv811wv (22.10.2018).

Kalantzsis M. and Cope B. 2013, *Multiliteracies in Education*, in Chapelle C.A. (ed.), *The Encyclopedia of Applied Linguistics*, Blackwell, Oxford, pp. 3963-3969.

Kendon A. 2004, *Gesture: Visible Action as Utterance*, Cambridge University Press, Cambridge.

Kress G. and van Leeuwen T. 1996, *Reading Images. The Grammar of Visual Design*, Routledge, London/New York.

Kozloff S. 2000, *Overhearing Film Dialogue*, University of California Press, Berkeley.

Lemke J.L. 1998, *Multiplying Meaning: Visual and Verbal Semiotics in Scientific Text*, in Martin J.R. and Veel R. (eds.), *Reading Science: Critical and Functional Perspectives on Discourses of Science*, Routledge, London, pp. 87-113.

Moskovich Y. and Shaf S. 2012, *Using Films as a Tool for Active Learning in Teaching*

Sociology, in “The Journal of Effective Teaching” 12 [1], pp. 53-63.
Norris S. 2004, *Analyzing Multimodal Interaction: A Methodological Framework*, Routledge, London.
O’Boyle E. and Sandonà L. 2014, *Teaching Business Ethics Through Popular Feature Films: An Experiential Approach*, in “Journal of Business Ethics” 121 [3], pp. 329-340.
O’Halloran K.L. (ed.) 2004, *Multimodal Discourse Analysis: Systemic Functional Perspectives*, Continuum, London.
O’Halloran K.L., Tan S. and Smith B.A. 2016, *Multimodal Approaches to English for Academic Purposes*, in Hyland K. and Shaw P. (eds.), *The Routledge Handbook of English for Academic Purposes*, Routledge, London, pp. 256-269.
Pandey S. 2012, *Using Popular Movies in Teaching Cross-cultural Management*, in “European Journal of Training and Development” 36 [2/3], pp. 329-350.
Querol-Julián M. 2011, *Evaluation in Discussion Sessions of Conference Paper Presentation: A Multimodal Approach*, LAP Lambert Academic Publishing, Saarbrücken.
Scollon R. and Levine P. (eds.) 2004, *Multimodal Discourse Analysis as the Confluence of Discourse and Technology*, Georgetown University Press, Washington, DC.
Sherman J. 2003, *Using Authentic Video in the Language Classroom*, CUP, Cambridge.
Timmis I. 2005, *Towards a Framework for Teaching Spoken Grammar*, in “ELT Journal” 59 [2], pp. 117-125.
Weinberg A., Fukawa-Connelly T. and Wiesner E. 2013, *Instructor Gestures in Proof-based Mathematics Lectures*, in Martinez M. and Castro Superfine A. (eds.), *Proceedings of the 35th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, University of Illinois at Chicago, Chicago, p. 1119.
Wildfeuer J. 2013, *Film Discourse Interpretation: Towards a New Paradigm for Multimodal Film Analysis*, Routledge, London.
Wittenburg P., Brugman H., Russel A., Klassmann A. and Sloetjes H. 2006, *ELAN: A professional framework for multimodality research*, in proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation. http://www.lrec-conf.org/proceedings/lrec2006/pdf/153_pdf (22.10.2018).

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 535-556
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p535

APPLYING SPECIALISED LINGUISTIC KNOWLEDGE IN THE CLASSROOM
ESP in social work discourse in Italy

JANE HELEN JOHNSON
UNIVERSITÀ DEGLI STUDI DI BOLOGNA

Abstract – The Social Work Corpus (SocWoC), a specially compiled repository of material of relevance to undergraduate trainee social workers consisting of almost 2 million tokens, has been presented in detail as regards compilation and description in previous work (Johnson 2016, 2017). Briefly, containing both spoken and written material in English, the corpus brings together service user interviews, case studies and other training materials, academic papers from the field of social work, newspaper articles on relevant issues, and government guidelines. Its validity is thus guaranteed since it features "actual examples of use which are drawn from the content area and which the learner is likely to have come across" (Flowerdew 1993, p. 239). Closer examination of the individual sub-corpora in SocWoC has enabled a more specifically linguistic focus on the world of social work, identifying a number of language patterns inherent to and typical of the different genres. These findings are further elaborated in this paper in order to develop classroom resources for use in an Italian undergraduate social work context in relation to the compilation of classroom material. The findings are also discussed in the light of examination of a small corpus of students' assessed essays (SocSEC) produced both before and after using SocWoC in the classroom. The focus is thus on the transmission of specialised linguistic knowledge from the corpus findings to practical application in the classroom, with considerations which may also be extended to other ESP disciplines.

Keywords: English for social work; corpus methods; classroom activities.

1. Introduction

Undergraduates following a course in English for Specific Purposes often find themselves faced with two challenges. On the one hand, they need to become aware of and confident with the various genres typical of their discipline, while at the same time they have to achieve this in a language which is not their first. We are specifically referring here to achievement of a familiarity with frequent words and phrases from the world of social work discourse, particularly those words and phrases which occur more often than might be expected when compared with a general text, as well as an awareness of whereabouts and how those words and phrases occur, in the sense of which words co-occur with them, what pragmatic function they might have, and in what sort of texts we might expect them to appear. Acquisition of such skills is an ongoing process and the teacher may play a major role in helping ESP students to discover the world in which they aspire to practise in another language so that they can operate successfully and achieve their goals (Dubois 2016, p. 26), enabling students to understand how and why language is used in the specific discourse community of social work and "apply their understanding proactively" (Solly 2008, p. 203).

We might say that this goal is typical of ESP in general, but it may be argued that it is particularly relevant to social services work, which depends largely on successful communication (Thompson 2010) thus requiring students to be exposed actively to such types of communication in order to acquire similar skills. Though specialised handbooks for social work are beginning to appear for students in the secondary education sector in Italy (e.g. Rudvin, Spinzi 2016), there is still little material for teaching specialised English to social work students at university level (Kornbeck 2003, 2008).

This article presents an outline of some of the salient linguistic features of a number of genres typical of social services and gives examples of how these may be exploited in the classroom. To do so, it draws on the contents of the Social Work Corpus (SocWoC) (Johnson 2016, 2017), a specially compiled repository of authentic material of relevance to undergraduate trainee social workers. Being a monitor corpus, it is constantly being amplified, with its contents at the time of writing including nearly 1.5 million words. SocWoC is made up of five sub-corpora, four of which feature mainly written material such as academic papers from the field of social work (ACAD), news reports on social work issues (NEWS), relevant newspaper opinion pieces (OPINION) and government guidelines (GOV). The fifth sub-corpus (MAT) consists of spoken material, and more specifically transcripts of service user interviews, case studies and other training materials. SocWoC thus represents examples of different genres in social work discourse (Flowerdew 1993).[1] Comparison between one sub-corpus and another enables a focus on both genre-specific differences as well as any differences in social work discourse at a macro level.

Containing authentic examples of different 'social work' genres, SocWoC may profitably be exploited in the classroom. The context here is a 30-hour English course at an Italian university for a class of about seventy second-year Social Work undergraduates, who have previously followed a

[1] A fuller description of an earlier version of SocWoC together with the rationale for text selection was given in Johnson (2017).

general English grammar course of 40 hours to reach an English language competence of B1+. The corpus findings may be used in such a class to raise awareness of formulaic language (e.g. Wray 2012), an important pedagogical issue since "learners tend to have a small inventory of formulaic sequences that they overuse and often have limited sensitivity to register differences" (Wray 2012, p. 235). Indeed, research (e.g. Rott 2009) has suggested that focus on formulaic sequences can improve native-likeness, while expanding students' vocabulary and lexical richness and variety for good, effective communication. Formulaic language includes recurring collocations and phraseology (Krausse 2008) and such a focus is particularly useful where certain phrases play a pragmatically specialised role (O'Keefe *et al* 2007, p. 63), as in social work communication.

A growing body of research has revealed the utility of learner corpora in material design for ESP/EAP courses (eg. Flowerdew 2001; Gilquin *et al* 2007). A small corpus of students' assessed essays (SocSEC)[2] was compiled to assess students' current language competence prior to classroom activity focussing on SocWoC corpus features. This learner corpus may thus serve as an initial point of reference from which to gauge current language awareness and identify areas of weakness in student competence. As the activities are introduced into the course, subsequent essays collected in SocSEC will act as benchmarks for assessing the utility of such activities.

The following section will introduce the corpus linguistics techniques used to query the corpus and introduce the type of classroom activities involved. Some of the shared features of social work discourse, as exemplified in the various sub-corpora, are then described in order to establish a 'core' set of lexical resources with which the student needs to become familiar. Similarities and differences between social work discourse and general English will be highlighted, while comparisons will also be made between the various SocWoC sub-corpora. Existing student language awareness of such lexical resources will be compared by means of reference to SocSEC and examples of practical activities used for further awareness-raising in the classroom presented.

2. Methodology

This section first describes the corpus tools used to extract the linguistic features of the SocWoC and the SocSEC corpora, suggests how needs

[2] At the time of writing, the Social work Students' Essay Collection (SocSEC) consists of approx. 100 short essays on topics such as Asylum seekers, Adopting Children, Homelessness, Parents and Marriage. Completed under examination conditions, it currently contains a total of 10,000 words.

analysis takes place on the basis of this latter corpus, and presents the type of classroom activities involved.

Corpus linguistics techniques have long been exploited in the ESP classroom (e.g. Boulton 2016, 2012; Biber *et al* 2010, p. 559). Much work has been done on how corpus linguistics allows certain features to be highlighted for subsequent close study and how this may be relevant to the language learner in general (e.g. O'Keeffe *et al* 2007; Adolphs 2006), as well as in an ESP context (e.g. Breeze 2015). Little, however, has been done as regards social work discourse. For this study, features of social work discourse were isolated using different types of corpus linguistics tools from a number of suites. Word lists, collocates, word clusters or 'lexical bundles' were extracted using Wordsmith Tools (Scott 2012), word profiles or 'sketches' using SketchEngine (Kilgarriff *et al* 2004), and lists of key semantic domains using WMatrix (Rayson 2008). Though there is little space here for giving a full description of how these tools work (see, for example, Adolphs 2006), more detail is given below as to how these queries help to exploit SocWoC.

Wordlists of the sub-corpora are examined to identify recurrent 'topics' in social work discourse, as well as frequently mentioned participants and this information may be used as a starting point for further investigation. Collocates of words occurring on either side of the node word can be used to investigate formulaic language. Similarly, lists of word clusters help to visualise the most common fixed patterns of a node word, enabling focus on phraseology, with the tools provided by the SketchEngine software enabling a focus on typical word usage within a phrase as well as a word's grammatical colligations. Analysis of semantic areas allows focus on the vocabulary which accumulates in particular semantic domains. Key semantic areas in the different sub-corpora are compared with those of general English text using WMatrix. Practical constraints such as time restrictions and lack of appropriate facilities are among the reasons why students are not asked to analyse the corpus themselves, which is why this article presents the author's analysis of some of the corpus data. However, Data-Driven Learning (Johns 1991) may profitably be implemented for future cohorts of students, indeed using corpus tools in the classroom is a tried and tested possibility in order to encourage autonomy and serendipitous learning (Bernardini 2000).

All the classroom activities for language awareness-raising involve the macro skills of reading, writing, speaking and listening. Beginning with the reading of selected concordances, the Tables below contain examples of possible classroom focus on language items via guided exercises such as intralingual transformation exercises using near-synonyms, and cloze exercises, as well as a focus on meaning by distinguishing differences between co-hyponyms, the matching of definitions with a word or phrase, creating taxonomies to show how topics are related, and translation, leading to broader discussion to compare the Italian context with that of the UK/US.

The subsequent section reports just some findings resulting from the corpus analysis, with suggestions for turning such findings into relevant classroom activities. It is worth stating here that this article is not intended to provide an exhaustive analysis of all the language features in SocWoC - a more detailed corpus analysis would certainly highlight further directions for developing classroom activities than there is space to describe here. Our aim is merely to draw attention to some frequent words and phrases in social work discourse and give examples of how these may be exploited in classroom activities to stimulate learning.

3. Corpus findings and classroom activities

3.1 Classroom focus (1): Wordlists – ‘care’

A simple wordlist enables focus on the most frequent words in a corpus. The specialisation of SocWoC is evident if the wordlist in the Annex is compared with that of the British National Corpus (henceforth BNC), with SocWoC featuring as many as twenty-six content words compared with eight in the BNC, with the content words all typical of social work discourse.

The SocWoC wordlist may be used to begin investigating certain issues in more detail. ‘*care*’, for example, is the 3rd most frequent content word in both SocWoC and SocSEC (the top 10 content words in SocSEC are *social, work, care, workers, people, practice, children, services, child* and *health*), but the words associated with ‘care’ in SocWoC and the learner corpus are different. For example, four out of five instances in SocSEC are preceded by TAKE, as part of the phrase *take care of*, which is overused here in relation to SocWoC (*take* is 42nd first left-hand collocate in SocWoC). This suggests that students need more exposure to uses of the word *care*.

The top first left-hand collocates of *care* in SocWoC differ according to the sub-corpus. Evaluative vocabulary such as *high-quality, safe, effective* and *inadequate* is frequently found in this position in the collection of government guidelines in GOV. In comparison, all the other sub-corpora, including the spoken MAT sub-corpus, more frequently have type rather than evaluation of care as premodifier, examples being *adult, community-based, end-of-life, foster, respite, alternative,* and *child*. Class discussion could first of all focus on possible reasons for this difference. Examination of expanded concordances show that it is due to the different communicative purposes of the texts involved. In class, students are then asked to focus more actively on the different collocations by reading selected extracts and sorting the collocates into type or evaluation of care, as in Table 1.

Read the extracts.[3] *The complexity of managing this patient group raises concerns that clinicians are struggling to provide appropriate* <u>*care*</u> *under the confines of a standard 10-minute consultation.* *The Triangle of Care for Dementia describes how meaningful involvement and inclusion of carers can lead to better* <u>*care*</u> *for people with dementia.* *The project aimed to provide a dedicated, sustainable, equitable Palliative* <u>*Care*</u> *Social Work model to adults and their carer(s)….* *We have seen some examples of outstanding care despite increased demand for services and challenging efficiency savings.* Sort the collocations of 'care' into: a) evaluation of care b) type of care

Table 1
Exercises with concordances of 'care'.

This exercise may be extended by getting students to match the definitions of the various types of care, suggesting possible equivalents in an Italian context, and focussing on the parameters used to describe the quality of care in different contexts, as in Table 2.

1) Match the definitions with these different types of care. e.g. palliative care, respite care, kinship care 2) Complete the gaps with a word or phrase describing the type of care 3) Read the descriptions below. How would you evaluate the care described? *effective, inadequate, high-quality* a) For example, following an inspection in June 2015 we cancelled the registration of a GP practice because inspectors had serious concerns about the service and the risks to people using it. During the inspection we identified one locum staff member who had treated patients but could not provide evidence that they were medically qualified to do so. The management of medicines was found to be unsafe and placed patients at serious risk of harm. [CQC_rep_001] [*inadequate*] b) Strong leadership with a culture of transparency where staff are valued for openly sharing concerns and reporting incidents or near misses. Clear lines of accountability and responsibility in all roles. Always putting patients first and working with other departments to maximise patient outcomes and experiences. [CQC_rep_001] [*high-quality*]
In groups, discuss and find Italian equivalents for these types of care: *respite care, kinship care, end-of-life care*

Table 2
Extension of exercises with concordances of 'care'.

[3] For reasons of space, only a small selection of concordances will be shown in the Tables to illustrate the type of activity. In the classroom, the students will have access to a number of carefully selected concordances. Possible answers and keys are shown within square brackets.

All these classroom activities are useful for raising awareness not only about vocabulary but also providing information and stimulating discussion about the differences between Italy and the UK as regards the provision of care.

3.2 Classroom focus (2): Wordlists – ‘children’

Many of the content words in both SocSEC and SocWoC are social actors and grammatical participants, e.g. CHILD/CHILDREN, FAMILY/FAMILIES, PERSON/PEOPLE, WORKER/S, STAFF, STUDENTS. While we might expect to find these in social work discourse in general, students need to be aware that these words might tend to associate with different words depending on the text type. By way of example, let us take the collocates of one of these participants – *children* – across the sub-corpora. This allows us to focus on formulaic chunks of language by investigating close collocates, such as the word/s immediately preceding our node word. In social work discourse, for example, is the word *children* closely associated with the same words as in everyday English? The top ten first left-hand content collocates of *children* in the British National Corpus and the written sub-corpora of SocWoC are shown in Table 3. The BNC’s frequent mention of handicapped and deaf children is possibly a sign of the age of the corpus itself, dating as it does from the beginning of the 1990s. Accepted terms have changed since then, and *handicapped* and *deaf* are very infrequent terms indeed in the SocWoC corpus, and never used in conjunction with children. This fact is itself worthy of further discussion in the classroom, leading to a discussion of sensitive terminology which may be picked up in later activities (such as in Classroom focus no. 3).

BNC	ACAD	GOV	NEWS	OPINION
young	safeguarding	safeguarding	protect	unaccompanied
their	(looked) after	authority	vulnerable	vulnerable
older	vulnerable	disabled	looked-after	looked-after
three	other	safeguard	poor	young
two	secure	individual	six	asylum-seeking
younger	young	(looked) after	Zealand	related
four	foster	older	abused	(looked) after
handicapped	abused	protect	five	help
small	adopted	protecting	young	refugee
deaf	protect	young	(looked) after	new

Table 3
Top ten first left-hand content collocates of *children*:
written SocWOC subcorpora and the BNC.

Throughout the written sub-corpora, one collocate is particularly frequent. This is *looked after*, also appearing in abbreviated form as *LAC* (looked-after children), or *CLA* (children looked after) (mainly in the ACAD subcorpus). The term itself never co-occurs with *children* in the BNC and thus we may presume that its acquired meaning of 'in care' is specific to social work. The meaning is not particularly transparent for the Italian social work student, unlike its near-synonym *children in care,* and thus needs particular practice in use.

Classroom work may focus on concordances to identify the meaning and typical phraseology of *looked-after*, contrasting it with the phrase *in care* and other related words such as *foster* and *adopt*. Some examples are:

(1) First, children who become looked after 'originate from the most disadvantaged social groups' BJSW 2015_1_3

(2) Since 2010, the government's annual SSDA903 data collection has included information about children who are looked after by local authorities, including the number of care orders made. BJSW 2016_4_8

(3) The campaign aims to tackle stigma around children and young people in care by interviewing inspiring care leavers who have gone on to have successful careers. OPI_GUA_3/10/16

(4) As there may be greater risks for looked after children and care leavers online, there may also be greater opportunities. BJSW 2016_4_13

(5) Emma stated that she does not want Chloe to go into foster-care at this time BJSW 2016_7_17

(6) only two out of the sixty children studied entered foster-care during the research

(7) The children placed with foster-carers ranged from birth to eighteen years, and were diverse in their nationalities and racial or ethnic heritage. BJSW 2016_6_19

(8) This is an order that gives the local authority permission to place the child for adoption. BJSW 2016_2_2

Word forms of *adopt* also collocate with *children* and belong to the same semantic field as *in care*, *looked after* and *foster*. Students may discuss the differences in meaning after reading the concordances: once a child is adopted s/he is no longer considered to be 'looked after' in the social work sense. As regards phraseology, reading the concordances enables colligating prepositions to be brought to the students' attention. For example, children are placed *in* [foster] care but *for* adoption. The different usage of prepositions may be highlighted by contrasting with the Italian equivalent *dare in adozione* (literally 'give in adoption').

Classroom activities could include compilation of a 'phraseology' table as shown in Table 4.

English language focus	***looked after***	***in care***	***foster***	***adopt***
Phraseology	*children + BE/BECOME+ looked after*	*children + BE + in care* *PLACE + children + in care*	GO INTO/ENTERfoster care PLACE in foster care PLACE with foster carers fostered children foster children	place child for adoption adopt a child
word order with *children*	pre-modified *looked-after children*	post modification only *children in care*	pre-modified: *foster/ed children*	adopted children
language variety?	UK English			
Italian equivalent?				

Table 4
Phraseology table for foster/ adopt/looked after + children.

As a follow-up activity students are asked to read the concordances and complete the exercises in Table 5 in class.

Read the concordances of *looked after, adopt, in care, foster*
1) Fill the gaps with the correct preposition (*in, for, into, with, to)*.
2) Complete the taxonomy: [*looked after* includes *foster/in care* but not *adopt]*
3) Rewrite the concordances of *looked after* replacing the phrase with *in care.*
4) Match the definitions below with the terms: ADOPT - FOSTER - LOOKED AFTER a) temporary custody or guardianship for children whose parents are dead or unable to look after them. [foster] b) the act of legally taking a child to be taken care of as your own [adopt] c) children living with relatives who aren't parents, e.g. children living with a grandparent/ older sibling [looked after]

Table 5
Follow-up activities for foster/ adopt/looked after + *children*.

Possible extension involves discussion of what is covered by the terms used for the Italian context, proposing a satisfactory equivalent in Italian for the phrases *foster*, *adopt*, *looked after*, and *in care*.

3.3 Classroom focus (3): Wordlists – 'people'

This sub-section will give examples of how another frequent participant in SocWoC provides scope for further classroom work. The most frequent first left-hand collocates show that people are often described in social work discourse as *young; older; disabled; LGBT; LGBTQI; vulnerable*; and *poor*. Of course, the head noun may also be post-modified and students can focus on the flexibility of post-modification – often by means of a prepositional phrase – beginning with a comparison of frequency in one sub-corpus. In the subcorpus of academic texts, for example, there are 68 occurrences of *disabled people* and 84 occurrences of *people with disabilities*. Students' attention can be brought to the fact that post-modification is more common than premodification in academic texts and they may be encouraged to compare this figure with the other sub-corpora. Variations in prepositional phrases may be identified in the classroom with reference to a list of selected concordances of the phrase people with, including a focus on co-hyponyms and their collocations, unpacking of acronyms specific to social work, and the grouping of adjectives according to meaning, as shown in Table 6.

Read the sentences containing the prepositional phrase *people with*.	Which phrases in column 3 do <u>not</u> refer to a medical condition? What do these acronyms stand for? ABI, TEC, LD, HIV, HCV. Which acronyms refer to medical conditions?
impairment/ disability/ Illness/ condition/ disorders/ needs are all general terms. They can be modified by type (e.g. *<u>learning</u> disability*) and/or by degree (e.g. *<u>severe</u> illness*). Find these general terms in the concordance examples and tick the box to indicate which adjectives of degree or type are associated.	
Complete the sentences with one of the following type modifiers: *dementia, eating, learning, mobility, ABI*	a. People with ....impairments find it difficult to move about. [mobility] b. Aspergers' syndrome is an example of a/n .... disability [learning] c. People with ....are included under the broad umbrella of 'adults with cognitive impairments' [ABI] d. people with ....may easily forget things [dementia] e. bingeing and vomiting are typical symptoms of a/n ....disorder [eating]
Group the adjectives of degree according to their meaning in the context of the phrase *people with x conditions*: (*severe, serious, acute, major, chronic, long term, long-standing, enduring, complex, multiple)*	very bad [*severe, serious acute,major*] something that doesn't go away [*chronic, long term, long standing, enduring*] complicated or involving more than one [*complex, multiple*]

Table 6
Classroom activities for *people with*.

Discussion in class could then focus on the topic of sensitive terminology (e.g. *disabled people* is a phrase less acceptable to the community nowadays than *people with disabilities*).

Students can also be made aware of the flexibility of post-modification which allows for further internal specification, such as "people with complex or severe disabilities including autism". The topic could be further extended by investigating the collocates of *disabilities*, *illness*, *conditions* and *health issues*.

3.4 Classroom focus (4): Clusters in social work discourse

While the separate words making up the clusters of all sub-corpora obviously feature in the single wordlist itself, viewed individually it is difficult to identify their function. A cluster wordlist makes this easier. The most frequent 3-word clusters in SocSEC are *a lot of*, *in order to*, *is very important*, *on account of*, *one of the*, *the most important*, and *the number of*. A comparison of the SocSEC list with the SocWoC list suggests overuse of simple quantity phrases in SocSEC (*a lot of*, *the number of*), while simple phrases such as *is very important*, and *the most important*, frequent in SocSEC, do not appear in SocWoC, which instead features more complex structural elements, such as *the importance of*. This would suggest more focus on these is required in the classroom. An examination of individual word cluster lists for each sub-corpus enables identification of similarities and differences. For example, some of the most frequent 3-5 word clusters in the sub-corpus ACAD are:

social work practice	*social workers and*	*one of the*	*in the context*
as well as	*social workers in*	*social work students*	*such as the*
the social work	*a number of*	*some of the*	*in the context of*
in order to	*the role of*	*there is a*	*social work in*
social work education	*in the UK*	*the impact of*	*the quality of*
in relation to	*the context of*	*that social workers*	*ways in which*
per cent of	*a range of*	*social workers to*	*social workers are*
the importance of	*part of the*	*the process of*	*for social workers*
in terms of	*of the social*	*a social worker*	*of the child*
social work and	*to social work*	*the need for*	*as part of*
for social work	*and social work*	*understanding of the*	*social works*
the development of	*the social worker*	*in the USA*	*the concept of*
the use of	*in this study*	*social work research*	*the fact that*

Table 7
Frequent 3-5 word clusters in the ACAD subcorpus.

With this list to hand, students can be asked to group phrases into quantity (*a number of, a range of, part of the, per cent of, some of the, one of the*), prepositional phrases marking relation (*as well as, in order to, in relation to, in terms of*), and abstract logical connectors (*the importance of, the fact that, the concept of, the context of, the role of, the impact of, the process of, the quality of, the use of, the development of*). A focus on structure highlights that these latter are often prepositional constructions. Prepositional phrases are examples of formulaic language that students need to memorize as chunks. The most frequent three-five word prepositional phrases in the academic papers sub-corpus ACAD are as follows:

the importance of	*the context of*	*the process of*	*a lack of*
the development of	*a range of*	*the need for*	*a sense of*
the relationship between	*the impact of*	*the quality of*	*the majority of*
the use of	*in order to*	*the concept of*	*as well as*
a number of	*in terms of*	*in relation to*	
the role of	*in this study*	*the nature of*	

Table 8
Frequent 3-5 word prepositional phrases in the ACAD subcorpus.

An example of classroom work focussing on such clusters in ACAD is shown in Table 9.

1) Read the list of word clusters. 2) Read the extended examples. Fill the spaces with the correct choice of Noun.	a) Social service and social workers, in particular, spoke about the.............. of resisting government agendas. One senior social service worker reflected the sentiments of many when she noted: "... we don't count our hours closely here. If we did, we'd probably all cry." [*importance*] b) Social workers in the research talked at length about the ....of their role and of the multifaceted issues that informed their work with sexually abused children. [*complexity*] c) Mother–infant foster placements are used where appropriate. A final decision about the ....for permanent removal is made in conjunction with the courts by the time the infant is six. [*need*]
The phrase *the process of* can be followed by what parts of speech?	a) We developed this case study by interviewing the group members about *the process of* working through their experiences of private pain and loss to addressing this in increasingly public ways. [gerund] b) In contrast, the existence of generations and the process of generational change can be conceived as *a* continuous *process of* renewal of participants in society. [nominal group]
Read the concordances.[4] Which other phrases follow the same pattern of phrase + Noun OR gerund? [e.g. *the importance of, a sense of*]	
Look at examples of the phrase *the development of.* Does it follow the same pattern? Focus on example (b) and explain the use of the –ing form here.	a) Training includes topics such as, for example, attachment theory, brain development research and the negative effects of cortisol on *the development of* behaviour and relationships. b) Notably, the social constructivist paradigm has a potential to confront the stigma attached to depressive experiences and can facilitate *the development of* helping relationships. c) The acquisition of a relevant and theoretically appropriate knowledge base thus constitutes a fundamental requirement for *the development of* competent social work practice with LGBT people.

Table 9
Examples of classroom activities for word clusters in academic texts.

[4] A list of concordances is given to students on a separate sheet.

*3.5 Classroom focus (5): Frequent clusters: '***the fact that***'*

As an example of how such a list may be further exploited in the classroom, let us look at the cluster *the fact that*, a cluster which occurs only once in SocSEC. Why should this cluster be important for non-native social work students? The utility of the 'factive nominal' (Lees 1960) for social workers is evident from Thompson's (1996) description: "expressing a proposition as a fact allows me to stand back from it and to comment on it as I bring it into my text, which means that if I want to comment on a proposition as I bring it in, I am more likely to express it as a fact" (1996, p. 210). We may see how useful it is in example 9 from the spoken MAT corpus:

(9) The fact that Ara's sexual abuse by her siblings as well as other family members occurred with the acknowledgment of her parents often kept me awake at night. MAT_CN_ISO_05

Labelling something as a fact, like packaging an issue by using nominalisation, makes it difficult to disagree with. Especially where the fact might be controversial, as in the above example, the utility of this frame becomes clear in social work communication. It also appears in other sub-corpora, clearly signalling writer/speaker evaluation, as in example 10:

(10) With the obvious impact of social care cuts on the NHS he is right to argue for increased resourcing of care - but isn't the real issue the fact that we have abandoned what was once a key element of our welfare state to what Behan terms "the market"? NEWS_GUA_17/10/16

A classroom task could be to define or paraphrase the 'fact' (the proposition) implied in the examples.

*3.6 Classroom focus (6): Frequent clusters: '***as well as***'*

Another frequent cluster in SocWoC is the linker *as well as*. This phrase would merit further attention in the classroom. It is underused as a coordinating item by Italian students who tend to stick to 'and' and is often used incorrectly, as shown in the single example in the learner corpus SocSEC.[5]

The different usage may be highlighted by giving students examples from the spoken MAT corpus, where the relative frequency of *as well as* is higher than in the other sub-corpora, as in examples 11-13.

[5] **"parents that fight and argue everyday can make children confused <u>as well as</u> a divorce".* [0075]*. A possible correct version might be; *"parents that fight and argue everyday can make children confused just <u>as much</u> a divorce can".*

(11) We spend a lot of time in staff meetings talking about the impact of culture on health care delivery. I sit on the hospital's diversity committee where we consider ways to serve the many cultural groups who populate this city as well as those newly arriving. bisman_cn_suz_01
(12) She conveys caring and compassion as well as respect for the service users and respect for the profession and role she performs. bisman_co5_cn_ter_01
(13) Well, she gets tired more easily so she's, uh, you know, she used to be really extrovert, but now she's just kind of really going back into herself. You know, not being able to go out much. It's really difficult for her to communicate as well as it's really affected her speech. woodcock_D _sue_01

Students could be asked to decide which of the examples of ‘as well as’ correspond to the meaning of ‘not only’ (examples 11 and 12), and which corresponds to the meaning ‘too’ (example 13). Follow-up may also include speaking and listening practice to focus on pronunciation of ‘as well as’ in a single chunk where relevant (as in examples 11 and 12). Students are then asked to paraphrase the excerpts to focus on the different usage, as in Table 10.

Read the sentences. Using a single clause, replace *not only/but also* with *as well as*	1. Not only does she convey caring and compassion but also respect for the service users and respect for the profession and role she performs. [=She conveys caring and compassion as well as respect for the service users and respect for the profession and role she performs.] 2. We consider ways to serve not only the many cultural groups who populate this city but also those newly arriving. [=we consider ways to serve the many cultural groups who populate this city as well as those newly arriving.]

Table 10
Classroom activities for *as well as*.

3.7 Classroom focus (7): Clusters in spoken social work discourse

Students need to be aware of clusters more frequent in one sub-genre rather than another, and this could be an opportunity for contrasting formal with more informal, or written versus spoken genres. The top fifty 3-5 word clusters in the spoken sub-corpus MAT are as follows:

what do you	*at the end*	*it sounds like*	*you like to*
I want to	*at the end of*	*like to do*	*you need to*
at the moment	*for you to*	*to do is*	*about how you*
I need to	*is that ok*	*to do that*	*be able to*
you want to	*the end of*	*to find out*	*have you got*
do you think	*to see you*	*to go back*	*I want to do*
I'd like to	*what do you think*	*to help you*	*if you are*
a lot of	*would you like*	*to talk to*	*need to do*
for you and	*do you know*	*want to do*	*need to find*
to talk about	*do you mean*	*we can do*	*need to talk*
we need to	*do you want*	*what do you mean*	*sounds like you*
you and your	*I come in*	*what I want*	
a social worker	*if you don't*	*what I want to*	

Table 11
The most frequent 3-5 word clusters in the MAT sub-corpus.

Hardly any of these appear in the top hundred SocWoC clusters as a whole so we may conclude they are typical of spoken and informal language. With the help of concordances showing the clusters in context, students can be asked to classify them according to their linguistic function, e.g. verbal groups (mental processes: *need, think, want, mean, find out, see, would like to*; verbal processes: *talk*; material processes: *do, go back, help, come*), interrogative forms, infinitives, time phrases (*at the moment, at the end of*), and generic quantification (*a lot of*).

Though some of these clusters are frequent in spoken language in general, a particular emphasis on mental processes and interrogative forms is even more typical of social work discourse. It is also important to look at the pragmatically specialised functions of phrases in social work discourse such as *sounds like you, is that ok, what do you mean*, which in context not only fulfill the function of politeness and face work (Brown, Levinson 1987) but also have the function specific to social work of 'reaching for feelings' (Shulman 2009), as in example 14.

(14) So, Karen, it sounds like you feel that you're left with the care of the children and you find it hard to get everyone up in the morning to get them to school.

In the classroom this could be approached by having students distinguish social worker turns from service user speaker turns in two stages: 1) prediction via discussion 2) verification by reading concordances, as in Table 12.

Who do you think says this phrase: the social worker or service user?	*or sounds like you don't think that anybody cares about how you are feeling* [social worker] *I feel like I've messed everything up.* [service user] *whether you are really feeling OK or whether you will feel like running away again* [social worker]
Listen to the dialogue and check whether you were right.	

Table 12
Classroom activities for speaker turns.

3.8 Classroom focus (8): Semantic fields

3.8.1 General discourse items

A comparison of the written sub-corpora with the MAT sub-corpus of mainly spoken text highlights a number of 'key' discourse segments:

yeah	*hello*	*um*	*thank_you*
you_know	*oh*	*I_mean*	*sorry*
yes	*do_you_think*	*right*	*they_say*
no	*obviously*	*I_see*	
all_right	*kind_of*	*uh*	

Table 13
Key discourse segments in the MAT sub-corpus.

Classroom work could focus on pragmatic meaning, position in the spoken turn, and language variety; for example, *kind of* is more frequent in spoken US material than in the UK texts. Another classroom activity would be to delete such phrases from the transcript and have students replace them appropriately.

3.8.2 The semantic field of 'worry'

Analysing semantic fields with the help of WMatrix (Rayson 2008) shows the semantic field of WORRY to be prominent in four out of the five sub-corpora (not present in ACAD). It is particularly prominent in the spoken sub-corpus MAT. For MAT for example, the components of the semantic field of 'WORRY' include the following lexical items:

> *worried, care, concerned, concerns, trouble, stress, worry, worrying, troubled, concern, nervous, bother, bothering, worries, troubles, stressful, anxiously, anxious, cares, distress*

Students can focus on the components and concordances of this semantic group in the different sub-corpora in order to highlight differences in phraseology, as in Table 14.

Read the concordances of WORRY /ANXIETY / CONCERN/ BOTHER/ CARE/ DISTRESS/ TROUBLE	which can be used as a verb? which is used as an adjective before the noun? which is used as an adjective after the noun? which is used as both? which describes a state? which describes a person? what does the adverb modify?
Now complete the sentences with the correct word form of the lemma in brackets	

Table 14
Classroom activities for items from the semantic field of WORRY.

Students may then be asked to focus on how usages vary according to genre. For example, here are some concordances of *concerned* in the spoken MAT sub-corpus:

to let you know that I am very **concerned** about what you said about hurting
Mrs. H, I 'm very **concerned** about your daughter. She is talking
Well , you know, I 'm **concerned** about making sure Ben's , you know
It's natural to be really **concerned** about your child's development
saving time because you're obviously **concerned** I don't really
You say you are **concerned** about Ben and different kinds of

Instead, concordances of *concerned* in GOV include the following:

we believe that the providers **concerned** have the ability and the capacity
to long-term improvement, we are **concerned** by the wide variation in the
these practitioners to be **concerned** about children's safety and wellbeing
carer, professionals were most **concerned** by the child's father who had a
For young children this often **concerned** emotional development linked to
of the young carer and adults **concerned**. If a local authority considers
records relating to the child **concerned** and any application under the
permission from the copyright holders **concerned**. To view this licence : visit
organisation , we are increasingly **concerned** about the crisis in social care as
coordinator by someone who was **concerned** about him. Jim had spent seven

A comparison would be necessary between the different uses of 'concerned' in GOV and MAT, which could present problems in understanding and reproduction for Italian native speakers. Classroom activities could thus involve students comparing meaning, usage, and possible translation, focusing on the different equivalents in Italian, such as *riguardare*, *coinvolgere*, and *preoccuparsi*.

4. Conclusion

This article sets out to show how a purpose-built corpus of social work discourse can provide material for awareness-raising as regards appropriate English language, through transformation of corpus evidence into classroom activities. Some examples of the language features of social work discourse extracted from SocWoC and its sub-corpora have been given, and suggestions provided for classroom exploitation.

Wordlists highlighted frequent words such as *care*, words which were then examined in greater detail to focus on collocates and usage/phraseology. The focus here has been on formulaic language, i.e. how words are combined in phrases, giving examples from authentic texts in the SocWoC corpus to stimulate students to form their own phrases. Wordlists have also enabled us to focus on the main participants in social work discourse such as children and people. Closer examination showed how longer phrases could become the subject of useful teaching in the ESP for social work classroom, particularly where comparison with their native Italian is not so helpful, as in the case of ‘looked after children’. Other examples included word clusters which were frequent in certain sub-corpora, including discourse markers having a certain structural form such as *as well as* and *the fact that*. Though these clusters are certainly not unique to social work discourse, they fulfill a specific pragmatic function here. Semantic vocabulary fields were also touched upon, including that of WORRY and related vocabulary such as *concern*. Differences between usage of *concern* in one sub-corpus and another were highlighted.

Exposure to authentic examples in context, particularly in the form of expanded concordances, provides starting points which can be exploited in the ESP language classroom in which students can become familiar with a range of social work language items. The macro skills of reading, writing, speaking and listening are brought into play through a variety of activities including cloze and transformation exercises, matching definitions and the creation of taxonomies. A focus on both meaning and form is given. Extension, where students apply their new language knowledge in discussion, takes place also by means of comparison with the local context.

The work-in-progress written learner corpus SocSEC has been referred to here in the first of its two roles: as an initial benchmark to see whether and to what extent students currently employ language typical of social work. Subsequently, evidence of successful language acquisition will be measured through adding further examples of students’ written work post-course to SocSEC. Consolidation of work on SocSEC will provide further needs analysis as well as post-course evaluation and this will be the focus of future study.

The SocWoC corpus itself is constantly being amplified in order to increase the wealth of language information available, particularly as regards examples of discourse from an English as a Lingua Franca environment, which is highly relevant to the social work context in Italy. Future research will need to focus on ways to exploit such new data, while at the same time, a greater variety of classroom activities will be developed in order to exploit the linguistic data already available.

With the increasing need to incorporate specialised discourse into English courses at university level (e.g. Breeze 2015), it is imperative for lecturers teaching English to become familiar themselves with the discourse world of their students and combine 'traditional' English language learning practices with elements from language for specific purposes. As illustrated in detail in this paper, one way to do this is to assemble an appropriate corpus of material, use corpus tools to extract language features, and apply language teaching practices to combine them with suitable classroom activities.

Bionote: Jane Helen Johnson is Research Assistant at the Department of Modern Languages, Literatures and Cultures of the University of Bologna. A qualified teacher of English as a Foreign Language, she currently teaches English for Social Workers on the Social Services degree course at the School of Political Sciences as well as English for Specific Purposes on the Masters degree course in Language, Society and Communication. Her research interests include English as a Medium of Instruction, Translation, Corpus Linguistics and Corpus Stylistics, and particularly Corpus-Assisted Discourse Studies (CADS), while her current research projects combine CADS and English for Specific Purposes.

Author's address: janehelen.johnson@unibo.it

Acknowledgements: We would like to thank the European Union for their precious financial support to the current research.

References

Adolphs S. 2006, *Introducing electronic text analysis. A practical guide for language and literary students*, Routledge, London/New York.

Bernardini S. 2000, *Systematising serendipity. Proposals for concordancing large corpora with language learners*, in Burnard L. and McEnery T. (eds.), *Rethinking Language pedagogy from a corpus perspective*, Peter Lang, Hamburg, pp. 225-234.

Biber D., Reppen R. and Friginal E. 2010, *Research in corpus linguistics*, in Kaplan R.R. (ed.), *The Oxford Handbook of Applied Linguistics*, Oxford University Press, Oxford, pp. 548-570.

Boulton A. 2012, *Corpus consultation for ESP: A review of empirical research*, in Boulton A., Carter-Thomas S. and Rowley-Jolivet E. (eds.), *Corpus-informed research and learning in ESP: Issues and applications*, Benjamins, Amsterdam, pp. 263-293.

Boulton A. 2016, *Integrating corpus tools and techniques in ESP courses*, in “ASp” 69, pp. 113-137.

Breeze R. 2015, *Teaching the Vocabulary of Legal Documents: A Corpus-Driven Approach*, in “ESP Today” 3 [1], pp. 44-63.

Brown P. and Levinson S.C. 1987, *Politeness. Some Universals in Language Usage*, Cambridge University Press, Cambridge.

Dubois M. 2016, *Science as vocation? Discipline, profession and impressionistic sociology*, in “Asp” 69, pp. 21-39.

Flowerdew J. 1993, *Concordancing as a tool in course design*, in “System” 21, pp. 231-244.

Flowerdew L. 2001, *Exploitation of small learner corpora in EAP materials design*, in Ghadessy M., Henry A. and Roseberry R.L. (eds.), *Small Corpus Studies and ELT*, Benjamins, Amsterdam, pp. 363-379.

Gilquin G., Granger S. and Paquot M. 2007, *Learner corpora: the missing link in EAP pedagogy*, in “Journal of English for Academic Purposes” 6, pp. 319-335.

Johns T. 1991, *Should you be persuaded: two samples of data-driven learning materials*, in Johns T. and King P. (eds.), *Classroom concordancing*, “English Language Research Journal” 4, pp. 1-16.

Johnson J.H. 2016, *Constructing an ESP course for Social Services undergraduates: corpus tools to the rescue*, paper presented within Teaching Practises in “ESP today”, Seminar 14, ESSE conference (English Society for the Study of English), Galway, Ireland.

Johnson J.H. 2017, *The SocWoC corpus: compiling and exploiting ESP material for undergraduate social workers*, in Sarré C. and Whyte S. (eds.), *New developments in ESP teaching and learning research*, Research-publishing.net, pp. 133-151. https://doi.org/10.14705/rpnet.2017.cssw2017.749 (05.05.19).

Kilgarriff A., Rychly P., Smrz P. and Tugwell D, 2004, *The Sketch Engine*, in Proc. EURALEX 2004, Universitè de Bretagne Sud, Lorient, pp. 105-116.

Kornbeck J. (ed.) 2003, *Language Teaching in the Social Work Curriculum*, Logophon Verlag, Mainz.

Kornbeck J. 2008, *A "Reverse Mission" Perspective on Second-Language Classes as part of Social Work Education Programmes*, in “Portularia: Revista de Trabajo Social” 8 [2], pp. 247-264.

Krausse S. 2008, *On the role of student research in the ESP classroom: A call for sustainable language skills*, in Fortanet-Gómez I. and Raisanen C.A. (eds.), *ESP in*

European Higher Education. Integrating language and content, Benjamins, Amsterdam/Philadelphia, pp. 233-243.
Lees R.B. 1960, *The grammar of English nominalisations,* Mouton, The Hague.
O'Keeffe A., McCarthy M. and Carter R. 2007, *From corpus to classroom*, Cambridge University Press, Cambridge.
Rayson P. 2008, *From key words to key semantic domains*, in "International Journal of Corpus Linguistics" 13 [4], pp. 519-549.
Rott S. 2009, *The effect of awareness-raising on the use of formulaic constructions*, in Corrigan R., Moravcsik E.A. and Ouali H.(eds.), *Formulaic Language Vol. 2: Acquisition, Loss, Psychological Reality and Functional Explanations*, Benjamins, Amsterdam, pp. 387-404.
Rudvin M. and Spinzi C.G. 2016, *Exploring Identities. Fundamentals in Human Sciences*, Trinity, Whitebridge.
Scott M. 2012, *WordSmith Tools version 6*, Stroud, Lexical Analysis Software.
Shulman L. 2009, *The skills of helping individuals, families, groups and communities*, 6th ed., Brooks/Cole Cengage Publishers, Belmont.
Solly M. 2008, *Implementing the Bologna process in Italy. A distinctive approach to language learning in domain-specific contexts*, in Fortanet-Gómez I. and Raisanen C.A. (eds.), *ESP in European Higher Education. Integrating language and content*, Benjamins, Amsterdam/Philadelphia, pp. 199-212.
Thompson G. 1996, *Introducing functional grammar*, Arnold, London.
Thompson N. 2010, *Theorizing Social Work Practice*, Palgrave Macmillan, Basingstoke.
Wray A. 2012, *What do we (think we) know about formulaic language? An evaluation of the current state of play*, in "Annual Review of Applied Linguistics" 32, pp. 231-254.

Annex

The top 100 words in SocWoC

1	the
2	and
3	#
4	of
5	to
6	in
7	a
8	social
9	that
10	for
11	is
12	with
13	as
14	on
15	work
16	this
17	be
18	their
19	are
20	it
21	was
22	care
23	by
24	or
25	from
26	not
27	they
28	have
29	at
30	i
31	were
32	an
33	workers
34	we
35	about
36	which
37	people
38	practice
39	children
40	more
41	can
42	has
43	services
44	but
45	child
46	who
47	al
48	et
49	these
50	there
51	all
52	you
53	also
54	health
55	what
56	been
57	research
58	may
59	had
60	how
61	other
62	support
63	such
64	her
65	one
66	service
67	them
68	when
69	some
70	family
71	will
72	between
73	do
74	she
75	need
76	within
77	so
78	p
79	should
80	our
81	than
82	time
83	s
84	professional
85	being
86	study
87	would
88	if
89	where
90	into
91	local
92	my
93	those
94	role
95	students
96	he
97	worker
98	mental
99	through
100	staff

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 557-579
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p557

CREATING LOW-TECH HIGH-RELEVANCE PERSONALIZED LANGUAGE CORPORA WITH SCIENCE POSTGRADUATES

CARMELA MARY WHITE
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Abstract – This paper proposes a technically simple way for students to exploit individual linguistically specialized corpora which are directly relevant to their professional learning curve. Established science researchers today use highly specialized forms of professional ESP communication in spoken, written, or mixed media, although the central vehicle for research validation remains the research article. Since academic success is measured in publications, postgraduates aspiring to a scientific career, while learning to design and perform research activities, must also gain communicative proficiency. Raising students' awareness of the linguistic characteristics and complex unwritten rules of research article construction can boost their communicative performance. Language acquisition theorists advocate the use of specialist corpora for acquiring specialist discourse competence, a theory supported by experimental studies involving the use and/or creation of personalized language corpora by students using concordancing software. However, science postgraduates can be reluctant to spend time learning to use even a simple concordance tool. Following a brief theoretical introduction and literature review, the author describes a teaching experience with a physics PhD group in which students created and studied personalized corpora of specialist research articles without a concordancing tool, whilst the teacher used concordancing to create tailor-made classroom activities based on the collective class corpus. Students prepared for peer-to-peer communication of research by analysing linguistic features, practising problematic points and comparing the rhetorical structure of their corpus articles with a generic model, thus building up a complex picture of how experts in their field do not simply convey information but work to persuade and reveal their stance.

Keywords: ESP; postgraduate EST course; personalized specialist field corpus; classroom activities.

1. Introduction

The overall aims of this paper are to explain why postgraduate science students need specialist language corpora and to identify and propose a technically simple method for getting students to create and use individual linguistically specialized corpora as an integral part of specialist English language courses. University administrators looking to improve efficiency and reduce costs ask why they should employ resources specifically for English for Science and Technology (EST) courses, rather than aim for general linguistic academic competency, to be taught in courses shared by all disciplines (Hyland 2009). To provide an insight as to *why* science students need purposely designed specialist EST courses, this introductory section first draws from the theoretical background in a variety of fields to explore both the nature of the target communicative competence and the relevance of specialist corpora to EST courses. This is followed by a short review of some recent papers to see *how* specialized text corpora have been successfully employed in ESP/EST teaching contexts, in doing so, a common obstacle reported by various authors is identified. The author's aim in the remaining sections of the paper is to suggest a possible solution that might encourage more EST practitioners to experiment with personalized corpora.

1.1. The need for purposely-designed specialist EST courses

Established hard science researchers use highly specialized forms of professional communication in spoken, written, or mixed media. Participants of a survey (White, forthcoming) carried out amongst teaching staff at the former Faculty of Science at Bari University report an increasing use of spoken English in their work and a variety of written genres, although the central vehicle for research validation remains the formal research article (RA) with more than 90% of survey participants declaring they publish exclusively in English. The academic pressure on individuals and institutions to perform well has never been greater, success being measured predominantly in terms of international publications (see, for example, the websites of the Italian ANVUR and equivalent UK REF, responsible for higher education institute research evaluation). Clearly, postgraduates aspiring to a career in science, while learning to design and perform research activities, must also gain communicative proficiency.

Hyland (2009) emphasizes the need for students to learn specialist discourses, but notes a tendency on the part of tutors and administrators to misrepresent academic literacy as a "single, overarching literacy which students have failed to master before they get to university", a shortcoming seen as easily remedied by "a few top-up English classes". This ideology, he warns,

> transforms literacy from a key area of academic practice, how we construct ourselves as credible linguists, psychologists or whatever, into a kind of add-on to the more serious activities of university life. (Hyland 2009, p. 9)

Hyland here touches on the important sociolinguistic aspects of academic literacy: unless students are given an opportunity to conceptualize the specific epistemological frameworks of academic research, they will have difficulty acquiring the necessary communication skills to grow professionally and enter their target community of practice. In community of practice terms (Wenger *et al*. 2002), postgraduate students are novices beginning initiation into a relatively exclusive community of practice, where knowledge of the community discourse plays a key role in community gatekeeping (Hyland 2009, p. 5). Thus, knowledge of specialist community discourse is an essential part of a postgraduate's target communicative competence.

1.2. Target communicative competence

EST teachers need to familiarize themselves with the cognitive frames of science to understand how the message to be conveyed and the chosen form of expression are linked. Tarantino recommends that applied linguists should, 'immerse themselves into the culture of the disciplinary community and become conversant with its languages and conventions' (Tarantino 2011, p. 180), since its language is inextricably bound to the conceptual framework both of scientific investigation in general and that of the specific discipline. She reminds us that neither everyday language nor even scientific outreach prose can be equated to scientific discourse because

> the content matter of specialist reports builds on different sources of knowledge, hence, it cannot be properly understood by people not trained in the specific field of research and application. Lay people can repeat technical terms in speech or writing. However, since they are unaware of the non-verbal dimensions which technical expressions embody, they lack the knowledge required to evaluate the reliability of a scientific text, criticise or expand its content. (Tarantino 2004, p. 84)

The complex scientific cognitive frames which science students must acquire have been developed over centuries, but for the purposes of this paper can be summarized in the following short tenets.

Firstly, scientists formulate and test measurable hypotheses about the world surrounding us. Early in their university careers, students learn how to produce qualitative and quantitative empirical data (mathematical studies, laboratory experiments) and begin the process of learning to write science (laboratory reports).

Secondly, the data thus produced are presented to support statements aimed at contributing to collectively accepted knowledge. The rationale behind the laboratory report format and content becomes clearer to students as they gradually acquire a scientific mindset, aided by a shift in reading from text books to research papers during their postgraduate studies.

Finally, every professional scientist acknowledges that the collectively accepted knowledge of his/her field, is always open to modification, through peer evaluation and further hypothesizing and testing, in direct contrast with the layman's tendency to view scientific knowledge as written in stone. These general scientific tenets model the language of the scientific discourse community as a whole, but "gaining specialized knowledge of a disciplinary community's genres and mastering them presupposes taking on the discipline's identity" (Dressen-Hammouda 2008, p. 234), so that students need to be exposed to specialist genres within their own discipline.

Numerous aspects of scientific discourse are described in the literature, including its grammatico-syntactical and lexical (Halliday 2004), rhetorical functional (Widdowson 1979, Swales, 1990) and metadiscoursal (Hyland, Tse 2004) characteristics and even the rationale behind its historical development (Gross 2002), which together set it apart from other forms of English. Halliday (2004, p. 162) identifies seven problematic grammatico-syntactical areas which are stumbling blocks for the uninitiated.

One of the most evident features is EST's tendency towards nominalization, which Halliday traces to the logical linear Theme-Rheme construction of arguments in research texts, consisting typically of a long sequence of connected steps. At any point the author may require a large number of previous concepts to support his/her next argument, therefore "the only way to package a piece of argument so that it becomes a natural Theme of a clause is to turn it into a nominal group" (Halliday 2004, p. 125).

The need to express complex new concepts thus leads to another feature of EST text: high lexical density. The following is a typical example characterized by a high lexical density and nominalization, coupled with a relatively simple clause structure.

> Among the most robust and sensitive <u>trace-gas detection techniques</u>, <u>quartz-enhanced photo-acoustic spectroscopy</u> (QEPAS) is capable of record sensitivities with a compact and relatively <u>low-cost acoustic detection module</u> (ADM). (Spagnolo *et al.* 2013, p. 1)

Another of the foremost characteristics of an EST text is its multimodality. Driven by the need to summarize bulk data efficiently and meaningfully, a considerable quantity of information is given through visuals, which have their own conventions and "grammar" (Kres, van Leewen 1996).

Furthermore, cutting edge science articles have long attracted the interest of linguistics for their lexical creativity (see Halliday 2004, Chapter 2) since new concepts require language to express them: the title of the paper cited above is a good example: "THz quantum cascade laser-based quartz enhanced photo-acoustic Sensor", which requires considerable "unpacking" or "decoding". The uninitiated would need each decoded element to be explained with at least one sentence to even begin to understand all the information conveyed to the expert reader.

The rhetorical conventions developed for each section of the standard Introduction Methods Results and Discussion (IMRAD) RA format form another layer of unwritten code. The various sections "perform different rhetorical functions and thus require different linguistic resources to realize those functions" (Swales 1990, p. 136), such as the need to juggle expertly past, present and perfect tenses in the Create A Research Space (CARS) model introduction section, the extensive use of the past simple passive in the method section, the use of epistemic modality in discussing results.

Metadiscoursal features are also frequent in EST texts, whether aimed at negotiating collective consensus, including hedging and other mechanisms to indicate stance, or helping the reader navigate the text. These structures, too, are "intimately linked to the norms and cultural expectations of particular cultural and professional communities" (Hyland, Tse 2004, p. 175). Thus, if our aspiring scientists are to succeed, they need to be aware of the conceptual frames, rhetorical structure and appropriate metadiscoursal features through which to share their work and gain acceptance within their specific target discourse community.

1.3. The case for using individual corpora in EST courses

We have seen how various studies suggest that raising students' awareness of the complexities of the unwritten rules can boost their communicative performance. Does work on corpora provide a suitable means of achieving this end?

An argument strongly in favour of using corpora directly with students is that corpora consist of authentic materials. Various language acquisition theories, including lexical bundles (Hyland 2008) and lexical priming (Hoey 2005), suggest that the use of authentic material in corpora provides excellent material for learners. Moreover, language theorists have further advocated the use of specialist corpora for acquiring specialist discourse competence. Indeed, specialist texts related to the student's own field of research surely fulfil Krashen's comprehensible input requirements (Krashen 1987) and those of the Vygotskian Zone of Proximal development (ZPD) theory (Lantolf, Poehner 2014). For a more detailed review of corpus-based language research and its application to language pedagogy see McEnery and Xiao (2010).

The literature provides a number of interesting experimental studies in ESP teaching contexts, using concordancing programmes with students and reporting positive outcomes in the main.

Lee and Swales report on a corpus-based EAP course for NNS doctoral students which they describe as an exercise in "technology-enhanced rhetorical consciousness-raising" (Lee, Swales 2006, p. 58). In this course, students compared their own writing with the expert texts. The authors discuss how corpora can be used in the classroom, concluding that the degree of specialization of the corpus used should depend on the level of disciplinary acculturation of the course participants. They also note that concordancing works better at lexico-grammatical and phraseological rather than structural levels (such as audience analysis and paper structure). This suggests that activities exploiting other methods of analysis should be developed alongside concordancing activities in an EST class, since the target communicative competence for PhD science students requires detailed knowledge of specialist RA structure.

Charles (2012) investigates the feasibility and value of an approach to teaching postgraduate EAP writing in which students construct and examine DIY discipline-specific corpora. She argues that the possible benefits outweigh the disadvantages which must characterize a comparatively diminutive corpus:

> First, it is unlikely that even a large general corpus will provide adequate data to respond to the highly discipline-specific queries of specialist students. Thus users may find that there are few or no examples, or that the examples retrieved are irrelevant, or even misleading. Second, I would suggest that the process of building their own corpus allows students to achieve deeper and more critical insights into the nature of corpus data itself. This understanding helps them to interpret corpus data more perceptively and to gain a greater appreciation of the pitfalls as well as the benefits of the approach. (Charles 2012, p. 94)

Milizia (2013), instead, introduced political science undergraduates to concordancing software using a specialist corpus. She observes that these students are interested in analysing political speeches, enthusiastic about working with authentic materials, and like to "get their hands on corpora and concordances themselves and find out about language patterning" (Milizia 2013, p. 158). In fact, students who completed the courses are reported in all three studies to have been enthusiastic.

However, unlike Milizia's undergraduates, who were both predisposed to language analysis and excited by their first real approach to research, only a handful of the postgraduates completed Lee and Swales' course, and although Charles finds most students were able to compile their own "quick and dirty" corpus, she reports that less than a third followed all the classes. Comments collected by the authors suggest that an important factor in the drop-out rate could have been the time factor: some students may have found the pay-off too high in terms of time/perceived benefit. Indeed, it is the experience of this author, when asking previous cohorts of science postgraduates in an anonymous course feedback questionnaire whether they would be interested in in learning to use a concordance programme, that few expressed a willingness to dedicate time to this activity.

This is corroborated by Adams' observations (2006), the purpose of whose research was to trial the concordancing approach in both discipline specific and general academic contexts. Adam remarks that hard sciences PhD students, who are already subject to time pressure, may be unwilling to dedicate time to learning to use a concordancing tool, adding that

> Software in this field is not usually designed for language learners to use: it is typically designed for language researchers and is often complex to operate. It also uses jargon familiar only to linguists and language specialists, adding to the time and effort necessary for students and teachers to learn. (Adams 2006, p. 6)

This points to an unresolved question for teachers of EST: might it be possible to take advantage of the benefits offered by individual specialist corpora in an EST context without the need for students to learn to use a concordancing tool?

In the light of the theoretical and pedagogical background of EST and the use of corpora discussed above, the author describes below a teaching experience aimed at exploring an alternative way to get individual students to create for themselves a technically simple, but linguistically specialized, language corpus. In Section 2 the author describes the constraints, overall aims and choice of materials for the course, while Section 3 gives details of the set-up of the individual and class corpora. Section 4 provides examples of teaching activities developed using the RA corpora, followed by brief reflections on the experience and possible implications for EST course development.

2. Course constraints and choice of materials

The activities which will be described in Section 3 were developed for a brief mandatory (16 hour in-class) language course offered each year by the physics department at Bari University to first year PhD students. The students who follow this course are generally highly motivated, though very pressed for time.

The data used in this paper refer specifically to the 32nd doctoral cycle course. Two of the 13 participants of the 32nd cycle were not native Italian speakers; in a brief written and spoken entry test one participant performed at the top of the CEFR level B1, eight in the B2 level range, and the remaining four at C1 level or over. A needs analysis revealed a wide variety and high level of complexity of the communicative tasks which these physics PhD students may be expected to perform in English during their PhD studies.[1]

[1] Including reading specialist RAs, lab manuals and emails, writing applications for stages abroad, preparing periodic research progress reports, or even collaboration writing RAs. Oral and/or mixed media tasks may include Skype interviews, meetings, poster preparation and conference poster

However, the time constraint necessarily confined the scope of the course to only a limited part of these tasks.

Subsequently, participants' own research was selected as the main focus, it was nonetheless agreed, together with the head of the Doctoral school, that the participants needed all round linguistic support. The following teaching strategies were therefore selected to work towards the overall goal of improving students' communicative performance in oral presentations supported with slides:

- class activities promoting fluency and practising delivery techniques and pronunciation;
- vocabulary expansion of specialist terminology and its collocations;
- vocabulary expansion of metadiscoursal chunks of both audience/reader – presenter/writer relations and speaker/writer's stance;
- exposure to authentic materials in the form of specialist RAs and audiovisual;
- material available on internet; the recreation (as far as possible) of conference conditions for the final presentations, with audience tasks and debriefing.

To provide authentic oral texts for the whole class, the author selected a number of audio and video materials freely available on the internet, ranging from undergraduate video lecture excerpts and podcasts to Doumont's (2010) excellent scientific presentations unit on the *Nature Scitable* website.[2] Such materials are very useful because they provide examples of generic scientific content and/or oral research presentations, yet they cannot provide examples of specialist discourse for every field. For the specific discourse of their individual fields, authentic specialist field materials therefore took the form of RAs whose content was directly relevant to the individual, thus providing the best available alternative to specialist oral text.

The reader may require a further explanation regarding the suitability of choosing *written* articles as material for an *oral* presentation course (beyond their highly relevant specialist discourse and content). Firstly, it should be remembered that an oral presentation is a multimodal form of communication, where the more formal register of written RAs, duly compacted, has its place in the slides. The course participants all had to produce slides, which of course

sessions, conference contributions, participation and notetaking in seminars and international summer schools, daily ELF communication with colleagues at home or in stages abroad.

[2] The complete ebook is available at https://www.nature.com/scitable/ebooks/english-communication-for-scientists-14053993, while the 3 videos can be viewed at:
https://www.nature.com/scitable/content/john-s-video-14024244,
https://www.nature.com/scitable/content/marie-s-video-14033591,
https://www.nature.com/scitable/content/jean-luc-s-video-14031500

requires the ability to produce suitable written sentences and subsequently to reduce them to bullet point format as needed. Secondly, as Hyland (2009) points out, conference presentations are more tightly organized and patterned than conversational speech. Indeed, differences between academic writing and speech should be seen on a continuum, rather than as clear-cut dichotomies. Thirdly, another important feature is shared by RAs and oral presentations: the interaction of text (whether spoken and/or written) and visuals. Hyland sums this concept up as follows:

> the complicating role of multimodal semiotics in spoken presentations of various kinds further undermin[es] a direct spoken-written split in communicative features. The non-verbal dimension of academic speech (and writing) in both constructing and conveying information of various kinds is substantial…. (Hyland 2009, p. 24)

Finally, the general RA format is followed more or less strictly in presentations (Mariotti 2012, p.67) and thus, the RA corpora provided important material for comparison, allowing similarities and differences in purpose, audience, structure, language and modality between the dominant RA, and the poster and presentation genres to be discussed in class.

A week before the course started, the 14 participants were asked to find two or more RAs in digital format, thus creating the basis of a personal corpus of specialist text. The criteria given to guide their choice were that the RAs should be of direct relevance to their own studies and written by acknowledged experts in their specialist field. Whether or not they had already studied them for the content was irrelevant.

The texts were vetted for suitability by the teacher. In some cases, students had selected reviews rather research articles, in which case they were asked to add another text presenting original research. A few texts had English language issues, and again the students were asked to add another text, preferably with at least one NS author, and warned that while the validity of the content and organization of the text were not in question and could therefore be evaluated, it could not be considered a reliable language model.

The resulting 28 articles formed the small class corpus of specialized text. Each student used both a digital and paper copy of their own texts. The purpose of this article is not to describe in detail the whole course or all the materials and activities, but to focus on the activities performed with the individual embryonic RA corpora and the class corpus, which are discussed in the next section.

3. Individual RA corpora and related class activities

The activities carried out using the RAs can be divided into individual, group or class exercises, or according to whether they are aimed at rhetorical-function- structural or grammatical-lexical analysis and practice. As stated above, the goal was to allow students to use their RAs as a corpus for rudimentary analysis without having to download and learn to use a concordancing tool. They either used built-in functions of standard pdf readers to analyse their individual corpora or performed exercises manually or orally.

3.1. Phase 1. Getting started with the individual RA corpora

Private study. As soon as their RAs had been vetted, students were instructed to familiarize themselves with their articles by:

- skim reading the abstract and conclusion and examining any visuals;
- reading the whole article, once without, then a second time with a dictionary, listing vocabulary which was new to them;
- using the pdf reader's 'Find' tool to check for more examples of the new vocabulary in their own texts.

Classroom activities. In pairs or small groups, students were asked to summarize briefly one of their chosen RA orally and explain in what way it was relevant to their own studies. They were then asked to discuss the following questions:

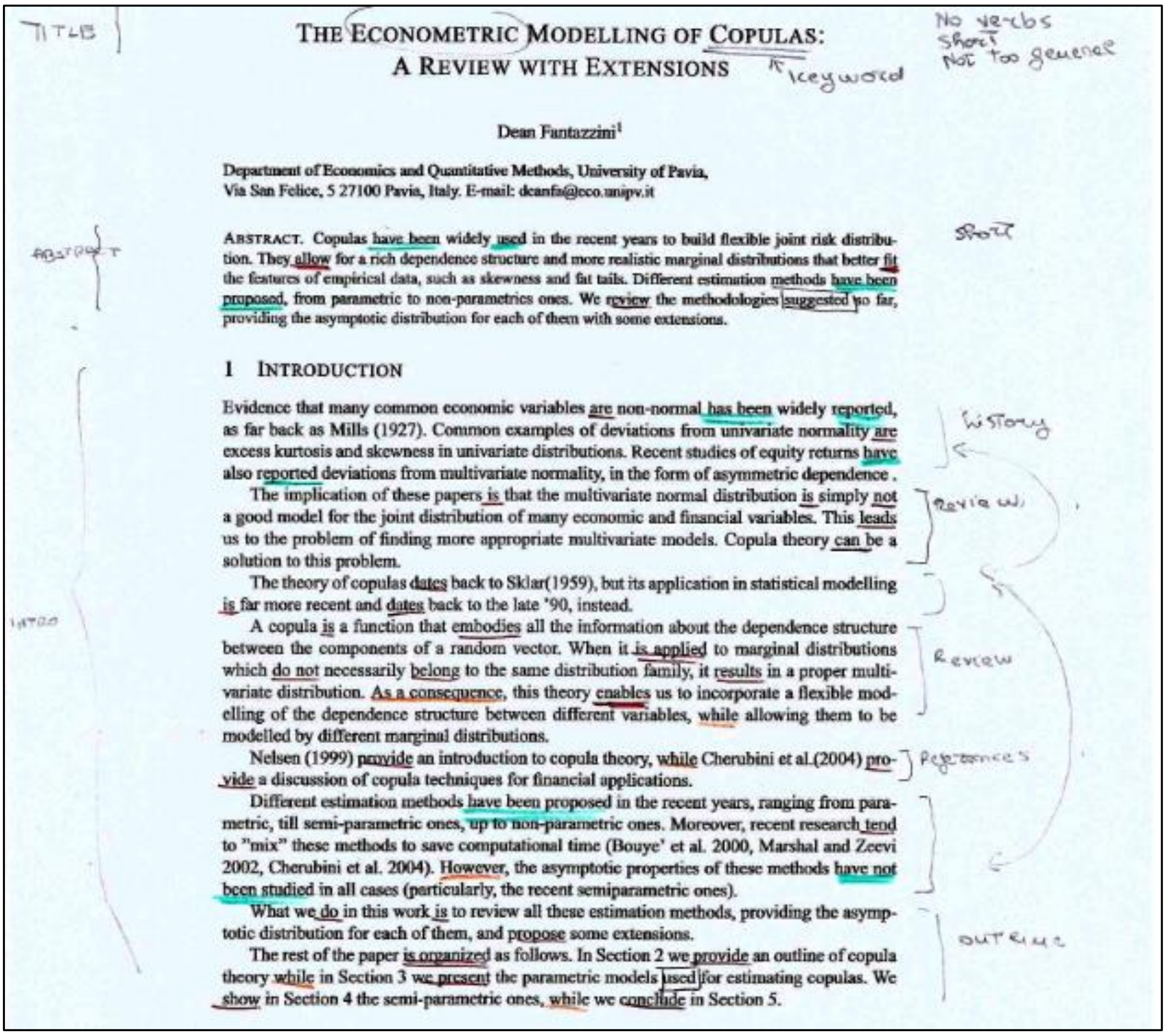
- Why do scientists need to write and present their research?
- Why do scientists need to write and present their research in English?
- What similarities and differences can you think of between a research article and a presentation? Consider content medium audience language organization.

This was followed by a class feedback session. A generalized model of the organization in sections of an RA was presented, which the students then quickly compared with their chosen texts, followed by further feedback.

3.2. Phase 2. Tense analysis of individual RA corpora

Private study. Students were first shown in class how to perform the tense analysis. They had the choice of doing this either manually using the paper copy, or digitally using their pdf reader (only possible with paid features of the software) or, possibly, by first laboriously moving the text into a Word file. Examples of both types were provided (Figures 1 and 2).

The analysis consisted in underlining each tense in a different colour, single underlining for active and double underlining for passive, to be completed for the abstract, introduction, and at least the first and final pages of every other section. Although this type of exercise may take some time, the coloured underlining makes the pattern of tense usage stand out clearly on the page, which is otherwise difficult to visualize. During the tense analysis, students were also encouraged to identify and mark reduced relative clauses, a high-frequency structure used in scientific prose to condense text.



THE ECONOMETRIC MODELLING OF COPULAS:
A REVIEW WITH EXTENSIONS

Dean Fantazzini[1]

Department of Economics and Quantitative Methods, University of Pavia,
Via San Felice, 5 27100 Pavia, Italy. E-mail: deanfa@eco.unipv.it

ABSTRACT. Copulas have been widely used in the recent years to build flexible joint risk distribution. They allow for a rich dependence structure and more realistic marginal distributions that better fit the features of empirical data, such as skewness and fat tails. Different estimation methods have been proposed, from parametric to non-parametrics ones. We review the methodologies suggested so far, providing the asymptotic distribution for each of them with some extensions.

1 INTRODUCTION

Evidence that many common economic variables are non-normal has been widely reported, as far back as Mills (1927). Common examples of deviations from univariate normality are excess kurtosis and skewness in univariate distributions. Recent studies of equity returns have also reported deviations from multivariate normality, in the form of asymmetric dependence .

The implication of these papers is that the multivariate normal distribution is simply not a good model for the joint distribution of many economic and financial variables. This leads us to the problem of finding more appropriate multivariate models. Copula theory can be a solution to this problem.

The theory of copulas dates back to Sklar(1959), but its application in statistical modelling is far more recent and dates back to the late '90, instead.

A copula is a function that embodies all the information about the dependence structure between the components of a random vector. When it is applied to marginal distributions which do not necessarily belong to the same distribution family, it results in a proper multivariate distribution. As a consequence, this theory enables us to incorporate a flexible modelling of the dependence structure between different variables, while allowing them to be modelled by different marginal distributions.

Nelsen (1999) provide an introduction to copula theory, while Cherubini et al.(2004) provide a discussion of copula techniques for financial applications.

Different estimation methods have been proposed in the recent years, ranging from parametric, till semi-parametric ones, up to non-parametric ones. Moreover, recent research tend to "mix" these methods to save computational time (Bouye' et al. 2000, Marshal and Zeevi 2002, Cherubini et al. 2004). However, the asymptotic properties of these methods have not been studied in all cases (particularly, the recent semiparametric ones).

What we do in this work is to review all these estimation methods, providing the asymptotic distribution for each of them, and propose some extensions.

The rest of the paper is organized as follows. In Section 2 we provide an outline of copula theory while in Section 3 we present the parametric models used for estimating copulas. We show in Section 4 the semi-parametric ones, while we conclude in Section 5.

Figure 1
Screen shot of a page of RA analysis with manual annotation produced by a student, provided to course participants as an example to follow.

Tense[1] analysis Example
key

present simple	past simple	present perfect	modal+ verb
present simple passive	past simple passive	present perfect passive	modal + passive verb

linking expressions   reduced relative clause

Preparation and characterisation of isotopically enriched[2] Ta_2O_5 targets for nuclear astrophysics studies

Abstract. The direct measurement of reaction cross-sections at astrophysical energies often requires the use of solid targets of known thickness, isotopic composition, and stoichiometry that are able to withstand high beam currents for extended periods of time. Here, we report on the production and characterisation of isotopically enriched[2] Ta2O5 targets for the study of proton-induced[2] reactions at the Laboratory for Underground Nuclear Astrophysics facility of the Laboratori Nazionali del Gran Sasso. The targets were prepared by anodisation of tantalum backings[3] in enriched[1] water (up to 66% in 17O and up to 96% in 18O). Special care was devoted to minimising[3] the presence of any contaminants that could induce unwanted[2] background reactions with the beam in the energy region of astrophysical interest. Results from target characterisation measurements are reported, and the conclusions for proton capture measurements with these targets are drawn.

1 Introduction

Stars spend most of their lives converting[3] hydrogen into helium through a sequence of reactions known[4] as the pp chain (for stars of mass *M* ≤ 1.5*M*_) or the CNO cycles (mainly for stars of mass *M* >1.5*M*_) [1,2]. An extensive experimental programme to study key reactions for hydrogen burning[5] in stars has been carried out over the last two decades at the Laboratory for Underground Nuclear Astrophysics (LUNA) facility in Italy [3,4]. Recently, the LUNA Collaboration has undertaken the study of proton-induced[2] reactions on 17O and 18O, important nucleosynthesis processes in several stellar sites, including[5] red giants, asymptotic giant branch (AGB) stars, massive stars, and classical novae [5,6]. In particular the ratio between the rates of the 17O(p, *α*)14N reaction (*Q* = 1191.8 keV) and the 17O(p, *γ*)18F reaction (*Q* = 5606.5 keV) affects the galactic abundance of 17O, the stellar production of the radioactive 18F nuclide, and the predicted[2] oxygen isotopic ratios in pre-solar grains [5]. For the study of such reactions, solid targets enriched[4] in 17O and 18O isotopes have been made.

[1] N.B. Be careful not to include words which are not active tenses! Remember many of the –ed and –ing forms are not active verb forms. The most common uses are: both –ed and –ing forms may be adjectives (2),–ing forms may be nouns (3), both –ed and –ing forms may introduce a reduced relative clause (4) and (5). Try to find examples of (4) and (5) in your texts.
[2] adjective
[3] noun form of the verb: when following a preposition, or when following a verb of pattern *verb + (do)ing*, or when it is the subject or complement of a verb. It may or may not be preceded by *a/an/the* etc. It may be plural (e.g. 'backings')
[4] reduced relative clause – past participle substitutes *which+ passive verb (any tense)*
[5] reduced relative clause – present participle substitutes *which+ active verb (any tense)*

Figure 2
Screen shot of a sample of digital tense analysis given to students as a model to follow.

3.3. Phase 3. Discovering how structural and rhetorical features of RA correlate with language choices

Classroom activities. The class was now ready to begin detailed discussion on the structure and discourse aspects of the RAs and to analyse the linguistic characteristics specific to each section. Using the same method as described above (alternating small group work and class feedback sessions), participants were guided through a series of concept questions to elicit a model for each section. They then compared the model with their individual corpora, noting structural and rhetorical aspects, such as evidence of the CARS introduction moves (Swales 1990) in the margins.

At least one two-hour session was dedicated to discussing each section and practising one or more related language points. By the end of this part of the course, working orally in small groups or pairs, the participants had explained the purpose of their work and a little of the theoretical background; they had described the method they had used in an experiment or a device they were involved in developing and they had described and interpreted a graph, or some other visual of their own research results. This was challenging work because although they all shared a common background knowledge of physics, they often knew little about each other's specialist fields.

4.Class corpus and related class activities

The 28 class corpus files were converted to .txt format by the teacher and loaded directly into AntConc (version 3.5.7; Anthony 2018), a freeware corpus analysis toolkit for concordancing and text analysis. The aim was to find examples illustrating difficult linguistic points chosen, as previously stated, on the basis of the author's professional experience. For the purposes of the course, time consuming "cleaning" of the files was deemed unnecessary. For this paper, two tricky words which often cause problems have been selected to illustrate the use that was made in class of the RA class corpus. The following paragraphs illustrate how activities were developed for illustrating and subsequently practicing an appropriate and correct use of "aim" and "allow".

4.1. Tricky word 1: "aim"

Clearly, as a word which expresses purpose, "aim" is most likely to be found in the introduction section of a RA, with possible use in the abstract and reiteration in the conclusion. These exercises and activities were therefore developed to use as part of the activities regarding the introduction section.

The difficulty many NNS find with this word is that they confuse the patterns which follow "aim" when it is used in the passive voice (meaning to be directed towards something), as an active verb (meaning to plan/aspire/intend to do something) and as a noun (meaning target, intended outcome). Halliday (2004) would probably recognize metaphor processes at work in most phrases using this word, the original meaning presumably being that of pointing a weapon.

A concordance run for “aim*” (Figure 3), executed on the 28 files of the class corpus found 27 hits, three of which were ignored because AIM was used as a proper noun (the title of a computer programme), and was therefore judged to be misleading for the purposes of the class exercise. The author then imported the hits into a Word file for ease of display, and analysed the examples, classifying them according to the various correct or incorrect patterns (the latter were all produced by authors who are presumably NNS), and reducing the number of characters before and after the target. This produced phrases which were both sufficiently detailed and would allow students to focus their attention on the target phrase (Table 1, Annex).

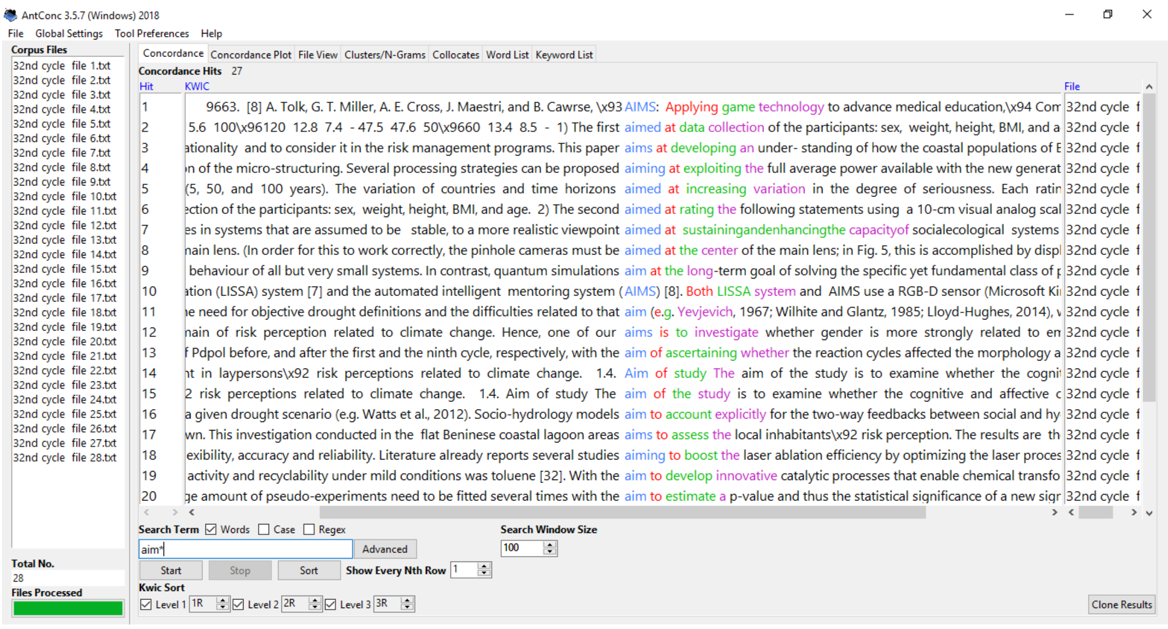


Figure 3
Screen shot of AntConc concordance results for “aim*”in the “dirty” 32nd cycle chosen RA class corpus. Hits 1, 10 and 27 were ignored because the word AIM was a proper noun. 5 examples with an incorrect pattern were found.

First the students were presented with (hopefully memorable) examples of the most common patterns. They were asked to imagine when or where they might hear/read the following and to paraphrase them:

- We aim to please, you aim too, please!
- Aim for the stars and you’ll reach the sky!
- The aim of the game is to cite your own name!

- Why are so many viruses aimed at Windows? It crashes just fine on its own, thank you!

Next, the class was given two or three correct examples taken from each group in the table, emphasizing the pattern. A written exercise followed, in which students had first to identify the 6 out of 12 further examples from the corpus with an incorrect pattern and then try to find a correct way of expressing the idea. Finally, in pairs they asked and answered the following questions:

- What is the aim of your PhD project?
- What are you currently doing towards your PhD? What is this aimed at?
- Are you aiming to do an International English certificate?

4.2. Tricky word 2: "allow"

In the experience of the author, the word "allow" is presented repeatedly to students in the passive expression "(not) [be] allowed to" in non-specialist EFL course books, but rarely in the active form. As online English exercises appear to reflect the same emphasis, this statement can be quickly tested by a rough and ready, although hardly scientific, method. If we insert "allow, English exercises" into an internet search engine, the results will almost certainly produce a strong preponderance of "allowed to" exercises, despite our not having specified this in the search. This is fine, it is correct English and presumably the idea of permission/authorization is the most common meaning of "allow" in general English corpora, since it is reiterated so frequently in EFL text books. However, science writers are more likely to need to explain mechanisms where one thing allows another to happen, expressing relationships between objects and processes.

Hoey's claim that "grammar is the product of the accumulation of all the lexical primings of an individual's lifetime" (Hoey 2005, p. 159) could explain why, in the author's experience, both students and professional NNS science writers alike have trouble with "allow". Somehow, the juxtaposition of "allowed" and "to", which students have used and heard so frequently and which sounds so similar to "allow to", may be priming them to extend the same juxtaposition to the active pattern, thus producing phrases with the active verb "allow" directly followed by an infinitive instead of the required noun/pronoun complement. Alternatively, it could simply be a cross-language priming from their native language. It would be an interesting point to follow up, but is beyond the scope of this paper.

The concordance run for "allow*" (Figure 4) produced 75 hits, which were examined and organized by the same method as the "aim" hits. Examples, of the 3 main correct patterns, and ones which were incorrect and the awkward solution with "one" as the complement, were chosen from among the hits and presented as “the Good, the Bad and the Ugly” (Figure 5). The students again worked on an exercise (Figure 6) in which some examples taken from the class corpus were correct and others required correction. In some cases, three alternatives were possible. It was necessary to add a few examples from the previous year because there were insufficient examples of the “allow + noun + to be/to be done” pattern in the 32nd cycle class corpus.

This exercise was followed by group work in which students challenged each other to describe different parts of a series of fantastical contraptions in Heath Robinson cartoons, such as The Deckeyrie For Unbalconied Flats[3] (Robinson, Browne 2015), answering the question: “what does this allow you to do?”.

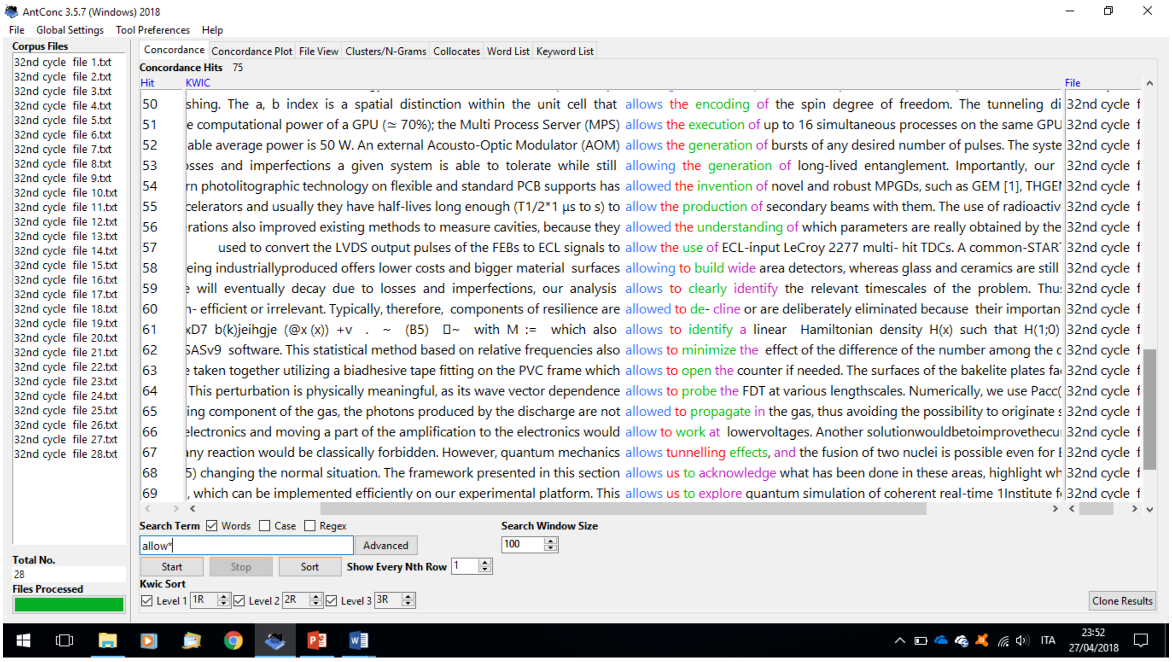


Figure 4
Screen shot of AntConc concordance results for “allow*” in the 32nd cycle chosen RA class corpus. Note hits number 60 and 65, the only two out of 75 hits to use the “(not) [be] allowed to do” pattern, and one used a similar pattern but in a reduced relative phrase. In the case of 65, “prevent” would probably have been preferable.

[3] Viewable at the following URL: https://www.spectator.co.uk/2017/05/there-was-method-to-the-madness-of-heath-robinsons-extraordinary-illustrations/. Accessed 27-04-2018.

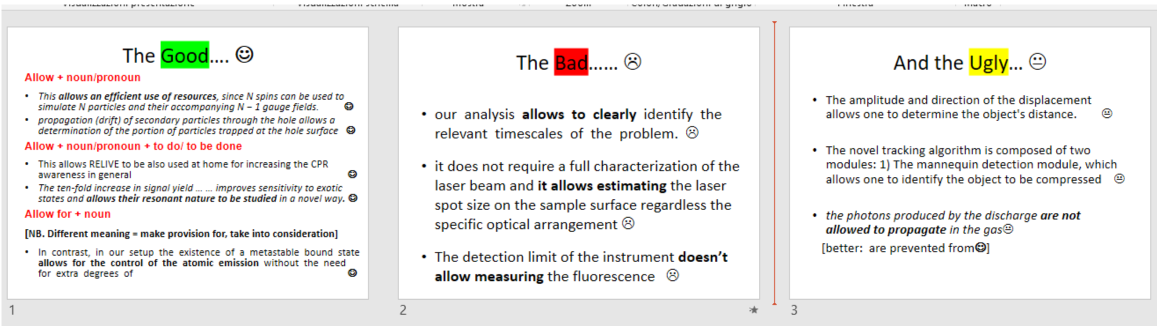


Figure 5
Screen shot of "allow" model and examples in ppt. slides shown to class during presentation phase of the exercise.

Exercise. Read the following extracts, most of which are from the class' chosen articles. In 5 of the examples the authors have used the verb correctly, in three it is correct but ugly, while in the remaining 5 they have got it wrong. Can you identify the 5 correct phrases? Now correct the others.

1. The hand detection module which allows one to allows one to identify and track the upper part of the rescuer\x92s overlapped...
2. it allows estimating the laser spot size on the sample surface regardless of the optical arrangement
3. Our platform allows direct measurements of the vacuum persistence amplitude and of the generated entanglement.
4. Brillouin microscopy can potentially allow more accurate determination of IOP
5. simple counting argument allows estimating the primordial 4He abundance
6. This perturbation is physically meaningful as its wave vector dependence allows to probe the FDT at various length scales.
7. the product of the two factors mentioned above allows the statistical uncertainty to be reduced.
8. III is devoted to the study of poles in the complex-energy plane, which allows us to extract crucial information relevant to the entanglement- by-relaxation protocol.
9. Furthermore, it introduces an important computational advantage, since it effectively allows one to split long simulations into a number of independent tranches, providing a natural and very efficient parallelization of the problem
10. the decrease of[sic] the time interval between consecutive pulses does not allow an efficient heat dissipation into the bulk material thus causing a temperature rise
11. This allows to reduce in a relevant way the amount of required memory in systems with multicomponent order parameters or in simulations of three-dimensional systems.
12. The panels are taken together utilizing a biadhesive tape fitting on the PVC frame which allows to open the counter if needed.
13. The amplitude and direction of the displacement allows one to determine the object's distance.

Figure 6
Screen shot of exercise discriminating between correct and incorrect phrases with "allow". The examples were taken from the class corpus, with the exception of a few examples from the previous year.

5. Conclusions

With this teaching project, the author set out to find ways of taking advantage of the benefits offered by individual specialist corpora in an EST context without asking the course participants to learn to use a concordancing tool. Acquiring greater awareness of, and ability to create, scientific discourse is crucial for young scientists in the first year of their PhD, which is an important transition stage between student and professional status. They have experience reading RAs but have not generally had the opportunity yet to consider in detail how a professional constructs a paper. The activities described above were integrated into the short course for first year physics PhD students with a view to offering students authentic, highly relevant materials as a resource for language learning in their specialist field. Student response to these activities was positive, the creation and use of personalized EST corpora helped to strengthen their motivation to dedicate time to language study because the language, rhetorical organization and content were perceived as highly relevant to their studies.

Some authors propose that students should learn to use a concordancing tool to become autonomous and responsible for their own continued progress (Lee, Swales 2006, p. 72). In the project presented here, instead, a choice was made to leave work with the concordancing tool kit exclusively in the hands of the teacher, given the tepid response of previous cohorts to the proposal of actually learning to use it themselves. This project has shown that it is feasible to use features of pdf readers in self-study activities to carry out rudimentary word searches in a limited specialist corpus relatively quickly, while simple manual highlighting on paper allow students to analyse linguistic aspects such as tense, modality, etc. Analysing overall RA structure and rhetorical functions lends itself better to work on paper, in the author's opinion, but could also be performed with margin comments in a pdf file. A pedagogical drawback remains that the implicit lack of commitment could mean that few will make the time to continue expanding their personal corpus after the course has finished. Clearly also, as some of the activities and classwork materials must be changed each year, it is a relatively time-consuming process for the teacher, as Charles (2012) also notes. Each individual teacher must decide if it is worth the effort.

A technical drawback of using a personalized corpus is that the amount of text is necessarily very limited, compared with large corpora. In fact, even using the class corpus, insufficient examples of some patterns were found. Nonetheless, the author found it was possible to integrate them with examples from previous years.

In planning a classroom project integrating work on corpora, EST teachers also need to take into consideration the impact of financial, bureaucratic and logistic aspects. Milizia (2013) describes giving students access to a large political corpus using WordSmith, although she does not mention if this is the freeware version, while Adams (2006) and Lee and Swales (2006) use both large and personalized corpora. Projects which involve access to precompiled corpora and/or paid software may have heavy costs (e.g. licenses to be made available to students and teachers, for use on university computers and/or at home) and involve bureaucratic and logistic challenges (e.g. time spent liaising with administrators and laboratory availability). Instead, the use of simple self-compiled corpora and freeware such as AntConc, as proposed here and in Charles' (2012) project, confer the advantage of ease of access both to the corpora and to the concordancing tools, thus reducing costs and simplifying bureaucratic/logistic issues.

Another unresolved issue in common with projects based on authentic discourse, raised here by the results of the concordancing of "aim" and "allow", is that of which linguistic model to choose to present to students. As can be seen from the percentage of errors found in this small class corpus consisting of highly specialized RAs, poorly written English sentences regularly slip through publishing vetting processes. The aim of this paper is not to address in depth the interesting question of whether NS models must always be preferred, or whether, considering that these students are using English as a Lingua Franca, we should accept any forms commonly used by NNS. However, we do need to point out to our students that at least three major aspects of their work will be evaluated when they submit an abstract or an article to a journal, or make a presentation: first and foremost its scientific content, but also their ability to express themselves in correct grammatical English and their knowledge of the unwritten rules of scientific discourse, a good command of which distinguishes an expert from a novice and will play an important role in helping them to pass the gatekeepers of their discipline.

In conclusion, as the author hopes to have shown, these three aspects are intimately interwoven and therefore need to be analysed and practised together, at specialist level. Much can be achieved using individually chosen specialist corpora in an EST class by a creative teacher, without asking students to learn to use concordancing software, thus providing aspiring scientists with exactly the ZPD material they need.

Bionote: Carmela M. White, 'CEL' and contract professor at Bari University, has worked in EST since 1985 and is currently attached to the Physics Department, where she is responsible for physics and materials science undergraduate and mathematics postgraduate courses, working closely with departmental staff to integrate language courses with the students' mainstream studies. She has presented several papers at the TESOL Italy annual conference on topics ranging from EST course and material development to motivation studies and tackling specific language difficulties faced by science students. The ideas presented here have been developed whilst teaching short courses for science PhD schools.

Author's address: carmelamary.white@uniba.it

References

Adams R. 2006, *Developing professional phraseology: a corpus linguistics approach*, in Mickan P., Petrescu I. and Timoney J. (eds.), *Social practices, pedagogy and language use: studies in socialisation*, Lythrum Press, Adelaide, pp 72-82.

Anthony L. 2018, AntConc (Version 3.5.7) [Computer Software], Waseda University, Tokyo, Japan. http://www.laurenceanthony.net/software

Charles M. 2012, *'Proper vocabulary and juicy collocations': EAP students evaluate do-it-yourself corpus-building*, in "English for Specific Purposes" 31 [2], pp. 93-102.

Dressen-Hammouda D. 2008, *From novice to disciplinary expert: Disciplinary identity and genre mastery*, in "English for Specific Purposes" 27 [2], pp. 233-252.

Doumont J. 2010 (ed.), *English Communication for Scientists*, NPG Education, Cambridge, MA. https://www.nature.com/scitable/ebooks/english-communication-for-scientists-14053993 (30.04.2018).

Gross A.G., Harmon J.E., Reidy M.S. 2002 *Communicating Science: The Scientific Article from the 17th Century to the Present*, Parlor Press, Anderson, SC.

Halliday M.A.K. 2004, *The Language of Science: Volume 5,* in Webster J.J. (ed.), *The collected works of M.A.K. Halliday*, Continuum, London.

Hoey M. 2005, *Lexical Priming: A new theory of words and language*, Routledge, London/New York.

Hyland K. 2008, *As can be seen: Lexical bundles and disciplinary variation*, in "English for Specific Purposes" 27, pp. 4-21.

Hyland K. 2009, *Academic Discourse,* Continuum, London.

Hyland K. and Tse P. 2004, *Metadiscourse in academic writing a reappraisal,* in "Applied Linguistics" 25 [2], pp. 156-177.

Krashen S.D. 1987, *Principles and practice in Second language Acquisition*, Pergamon Press Inc., New York.

Kress G. and van Leewen T. 1996, *Reading Images. The Grammar of Visual Design*, Routledge, London/New York.

Lantolf J.P. and Poehner M.E. 2014, *Sociocultural Theory and the Pedagogical Imperative in L2 Education: Vygotskian Praxis and the Research/Practice Divide*, Routledge, London/New York.

Lee D. and Swales J. 2006, *A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora*, in "English for Specific Purposes" 25 [1], pp. 56-75.

Mariotti C. 2012, *Genre Variation in Academic Spoken English: the Case of Lectures and Research Conference Presentations*, in Maci S.M. and Sala M. (eds), *Genre Variation in Academic Communication: CERLIS Series Volume 1*, CELSB Libreria Universitaria, Bergamo, pp. 63-84.

Milizia D. 2013, *Phrasal Verbs and Phrasal Units: Political Corpora within the Walls of the Classroom*, in Desoutter C., Heller D. and Sala M. (eds.), *Corpora in specialized communication: CERLIS Series Volume 4*, CELSB Libreria Universitaria, Bergamo, pp. 135-164.

McEnery T. and Xiao R. 2010, *What corpora can offer in language teaching and learning*, in Hinkel E. (ed.), *Handbook of Research in Second Language Teaching and Learning* Vol. 2, Routledge London/New York, pp. 364-380.

Robinson W.H. and Browne K.R.G. 2015, *How to Live in a Flat*, Bodleian Library, Oxford (first published 1936).

Spagnolo V. et al. 2013, *THz quantum cascade laser-based quartz enhanced photo-acoustic sensor*, in “38th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)”, pp. 1-3.
Swales J. 1990, *Genre Analysis*, Cambridge University Press, Cambridge.
Tarantino M. 2004, *Epistemic and dialectic pathway to knowledge, meaning and language Advancement*, in “LSP & Professional Communication” 4 [1], pp. 69-88.
Tarantino M. 2011, *Modality categories in multimedia genres*, in “Rassegna Italiana di Linguistica Applicata” 3, pp. 157-183.
Wenger E., McDermott R. and Snyder W.M. 2002, *Cultivating Communities of Practice: A Guide to Managing Knowledge,* Harvard Business School Press, Boston, MA.
White C.M., forthcoming, *EST or CLIL? Integrating language courses within science degree programmes,* accepted for publication in “PERSPECTIVES A Journal of TESOL Italy”.
Widdowson H.G. 1979, *Explorations in Applied Linguistics*, Oxford University Press, Oxford.

Annexes

Hit n°	Beginning of phrase	Target and completion of phrase	Text N°
	a. "Aim" used correctly as noun (+ of), or as noun [be] + infinitive		
21	the system was adopted for two different	aims: 1) to estimate the accuracy of the hand tracking	24
11	the difficulties related to that	aim, which we will not repeat here	18
14	1.4.	Aim of study.	8
15	The	aim of the study is to examine whether the cognitive and affective components of risk	28
13	with the	aim of ascertaining whether the reaction cycles affected the morphology and the dispersion	3
12	Hence, one of our	aims is to investigate whether gender is more strongly related to emotional reactions	28
	b. Verb "aim" use correctly with at + noun (hit 9 only), or passive verb [be] aimed at +-ing		
9	In contrast, quantum simulations	aim at the long-term goal of solving the specific yet fundamental class of problems	19
8	the pinhole cameras must be	aimed at the center of the main lens	4
2	1) The first	aimed at data collection of the participants	24
6	2) The second	aimed at rating the following statements using a 10-cm visual analog scale	24
7	a more realistic viewpoint	aimed at sustaining and enhancing the capacity of social ecological systems to adapt to uncertainty	26
5	The variation of countries and time horizons	aimed at increasing variation in the degree of seriousness.	28
c. Verb "aim" used correctly active + infinitive			
18	Literature already reports several studies	aiming to boost the laser ablation efficiency by optimizing the laser process parameters	8
25	resulting in long-term adaptations	aiming to reduce impacts of drought in the future	18
16	Socio-hydrology models	aim to account explicitly for the two-way feedbacks between social and hydrological processes	18
17	This investigation conducted in the flat Beninese coastal lagoon areas	aims to assess the local inhabitants risk perception.	27
22	great potential for use in drought research is comparative analysis, which	aims to find patterns by analysing a large set of catchments with a wide range	18
23	RELIVE	aims to go beyond the state of the art accurately estimating the CC depth	24
d. "aim" used incorrectly			
19	With the	aim to develop innovative catalytic processes that enable chemical transformations to be performed	3
20	pseudo-experiments need to be fitted several times with the	aim to estimate a p-value and thus the statistical significance of a new signal	11
24	to study the entire interrelated system with the	aim to put the pieces of the puzzle together.	18

26	focus on a specific point or research question with the	aim to solve part of the puzzle, or to study the entire interrelated system	18
4	Several processing strategies can be proposed	aiming at exploiting the full average power available with the new generation of ultrafast laser	8
3	This paper	aims at developing an understanding of how the coastal populations of Benin perceive natural	27

Table 1
24 hits of "aim" organised according to pattern, including 6 incorrect examples.

Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 581-597
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p581

METHODOLOGICAL PROPOSAL TO BUILD A CORPUS-BASED ONTOLOGY IN TERMINOLOGY

ISABEL DURÁN-MUÑOZ
UNIVERSIDAD DE CÓRDOBA

Abstract – Corpora are an indispensable resource to improve quality both in the linguistic and conceptual dimension of terminological projects. However, while there is complete agreement that specialised corpora are vital in the linguistic dimension of any terminological project (e.g. to select real contextual examples), there are three different approaches with regard to the conceptual dimension and not all of them employ corpora in their projects. In an attempt to shed some light on the advantages that corpora bring to the representation of specialised knowledge in terminology, this research follows the ontoterminography methodology (Durán-Muñoz 2012) to propose the building of a corpus-based ontology within a terminological project, in particular a specialised resource about an adventure activity (canyoning) in English. More specifically, it describes the different steps that are required to create such an ontology, from the analysis of the specialised domain and the compilation of the corpus to the representation of the specialised knowledge in the form of a corpus-based ontology.

Keywords: specialised discourse; specialised corpus; corpus-based ontology; ontoterminography; terminology.

1. Introduction

Corpus Linguistics has been one of the linguistic disciplines that has been most influential in Terminology. So much so that nowadays the idea of employing specialised corpora in terminological projects is fully accepted in the community of terminologists, as corpus has become an indispensable resource for any work of this nature (Dash 2000, p. 27). The use of corpora has, therefore, become an essential tool for current terminological projects, including production of specialised dictionaries and databases, giving rise to what Leech (1992, p. 106) considered “a new way of thinking about language”. Corpora can be used at any stage of a terminological project, from the selection of contexts to be included in the terminological resources to the study of concepts and the conceptual relations among them to create knowledge representation and become familiar with the specific subject of their work, what Cabré Castellví (1999, p. 144) calls *cognitive competence*.

While the usage of specialised corpora in the linguistic dimension in terminology is now unquestionable, no clear consensus exists over their use in the conceptual dimension. There are some methodologies that employ specialised corpora in order to analyse concepts and conceptual relations in a specialised domain by means of the study of their designations (or terms) in context (see L'Homme 2008). In other words, these methodologies follow a bottom-up approach and consider the text (or corpus) as the starting point to reach the specialised knowledge. On the other hand, we also find methodologies that follow a top-down approach, in accordance with theoretical principles of the traditional terminology (see ISO 704 2000, ISO 1087-1 2000; Roche *et al*. 2009), and analyse specialised domains by means of the initial identification of concepts by domain experts. Hence, they follow an onomasiological approach and analyse specialised domains from concept to term. Finally, there is another type of methodology that follows a middle-out approach in order to examine the specialised domain, which combines both views: it starts from the corpus in order to extract frequent terms and gain knowledge about a particular domain, but, at the same time, it requires the assistance of domain experts to build the conceptual representation of the specialised domain (Temmerman, Kerreman 2003; Faber *et al*. 2005; Durán-Muñoz 2012).

This research revolves around this conceptual dimension and aims to contribute to this middle-out approach, which combines specialised corpora and expert knowledge, by illustrating the creation of a corpus-based ontology for a particular specialised domain: the adventure activity of canyoning. To do so, it will follow the ontoterminography methodology (Durán-Muñoz 2012), a middle-out approach inspired by Frame-based Terminology (Faber *et al*. 2005) and the theoretical principles of the Sociocognitive Theory of Terminology (Temmerman 2000).

The remaining part of the paper is structured as follows: Section 2 outlines the concept of ontology and its fruitful relationship with Terminology; Section 3 describes the methodology followed and the different steps: 1. Corpus compilation, 2. Extraction of candidate terms and conceptual relations; 3. Expert consultation, and 4. Creation of corpus-based ontology. Finally, Section 4 presents some concluding remarks and future work.

2. Ontologies and terminologies, a fruitful relation

Ontologies have aroused the interest of researchers in different fields, such as Terminology, Artificial Intelligence or Computational Linguistics, and ontology-oriented applications are commonly found as part of information systems, databases, natural language processing systems, knowledge based systems, etc. As Roche (2007, p. 47) states,

> Such a success is mainly due to what an ontology promises; it means a way of capturing and representing a shared understanding of a domain that can be understood and used by humans as well as by software.

The term *ontology* has its origins in the field of philosophy and it is considered a branch of metaphysics that deals with the nature and the organisation of reality. However, it has gained special relevance in the field of Artificial Intelligence, Knowledge Engineering, and Computational Linguistics in recent decades as a means of modelling knowledge, and it has acquired a more pragmatic and applied meaning derived from its original meaning.

In this computing-related field, the term *ontology* has numerous definitions, but the most accepted and cited by authors is the one proposed by Gruber (1993, p. 199): "An ontology is a formal, explicit specification of a shared conceptualisation". This definition, already considered as standard, includes the most relevant aspects of an ontology and its basic principles, namely:

- The term *conceptualisation* refers to an abstract model of the domain, or some phenomena of the world that it represents, which is intended to identify the most relevant concepts.
- The term *explicit* indicates that the concepts that constitute an ontology are defined in an unequivocal way.
- The adjective *formal* refers to the fact that the ontology must be expressed in some form of computer readable language by means of an identical formalism, so that it can be reused and understood by any machine regardless of the place, the platform or the language of the computer system that uses it.
- *Shared* reflects the notion that an ontology represents knowledge shared and accepted by the group or epistemological community to which it refers, and not only that of an individual.

As observed, *ontology*, also named *conceptualisation*, refers to a set of hierarchically organised concepts represented in a computer system whose purpose is to support various applications that require specific knowledge on the subject that ontology represents (Moreno Ortiz 2004, p. 31). In the same vein, Roche (2007, p. 47) summarises: "an ontology is a shared description of concepts and relationships of a domain expressed in a computer readable language."

This latter meaning of the term *ontology* is gaining particular significance in the field of Terminology as it can be considered a bridge between knowledge representation in the mind and language. In Faber *et al*.'s (2005) words:

> Terminology is not only a matter of terms and term entries that endeavour, no matter how imperfectly, to represent slices of objective reality. In this sense, the representation of a specialized field should be more than a list or even a configuration of objects linguistically translated into either simple or compound nominal forms. It is necessary to situate concepts in a particular setting and within the context of dynamic processes that define and describe the principal event in the specialized field in question.

We are witnessing a great increase in the use of ontologies[1] to carry out research as well as to produce terminological resources, such as specialised dictionaries or databases, in an attempt to organise specialised knowledge (concepts and relations among them) in a clear and systematic way by means of computer applications. An example of this are the numerous conferences that focus on this line of research, to name a few: *Terminology and Artificial Intelligence* (TIA), *Terminology and Knowledge Engineering* (TKE), *ToTH: Terminology & Ontology: Theories and applications*, among others, which confirm the approach and cooperation between Terminology and Knowledge Engineering.

This relationship has led to the fruitful development of ontologies and computer applications for the management of specialised knowledge in the field of Terminology, which has facilitated a great advance in ontoterminography. In this context, specialised corpora are crucial to providing both conceptual and linguistic information and are currently the starting point of any terminological project aimed at producing specialised resources (dictionaries or databases), as we will see in the next section.

[1] In modern terminology we can see that, instead of the term *ontology*, other terms are employed that show similar features: knowledge representation, semantic categorisation or semantic frame.

3. Methodology: building a corpus-based ontology

As stated above, this research follows the ontoterminography methodology (Durán-Muñoz 2012) to propose the production of a corpus-based ontology in a terminological project. This methodology presents a middle-out approach and is inspired by the theoretical principles of the Sociocognitive Theory of Terminology (2000) and the Frame-based Terminology (*Faber et al*. 2005), a theory which employs semantic frames[2] as a basis for structuring expertise and creating separate representations of the working language. By way of an example, this paper depicts the different steps that are required to create such an ontology, from the compilation of the corpus to the representation of the specialised knowledge in a specialised resource about the adventure activity canyoning in English.

The ontoterminography methodology divides the terminological project in six main phases: 1) Design of the project and analysis of the specialised domain; 2) Compilation of the specialised corpus; 3) Ontoterminography management; 4) Elaboration of the ontoterminography database; 5) Validation, and 6) Edition of the terminological product (database, dictionary, etc.). The order established here does not imply that a strict sequence must be followed, as some tasks can overlap and occur simultaneously.

In view of the focus of the paper, only the first three steps will be considered, paying particular attention to corpus compilation. A specific adventure activity, canyoning, will serve to illustrate the building of such a corpus-based ontology.

3.1. Design of the project and analysis of the domain

This first step refers to the preparatory work of any terminological project and involves a series of decisions that serve as the basis for the entire process. In this initial phase, it is necessary to establish the pragma-linguistic variables that characterise the final resource and that need to be maintained throughout the entire process, always keeping in mind the

[2] Fillmore (1968) was one of the first to introduce the concept of *frame* in linguistics, considering it as a linguistic tool to represent extralinguistic events. Atkins and Rundell (2008: 145) explicitly define what a semantic frame is: "A semantic frame is a schematic representation of a situation type (e.g. speaking, eating, judging, moving, comparing, etc. - activities and situation which make up our everyday life) together with a list of the typical participations, props, and concepts that are to be found in such a situation; these are the semantic roles, or 'frame elements' (FEs)."

objectives of the research. These variables refer to the topic, the languages of the resource, the target users, the function to be covered, as well as the human and technical resources required. Besides the team working on the project, it is crucial to select domain experts in the working languages that assist terminologists during the whole process, but particularly during the conceptual representation of the domain (top-down approach).

Once the project has been designed, the analysis of the domain comes next. This consists of the study of the domain in a broad way and from different perspectives: socio-economic, pragma-linguistic and even a translation approach (if multilingual), so that terminologists can acquire broad knowledge of the domain and of the possible needs and difficulties of the project. To fulfil this stage, terminologists should consult domain experts[3], specialised journals, reports on the situation, relevant entities in the domain, existing legislation (if applicable), etc. It is also important to carry out an assessment of the terminological resources available at that time for that particular specialised domain. The analysis of available resources will allow terminologists to undertake an *in vitro* study of the terminology and learn from the definitions and other information included in those resources.

At this stage, terminologists have acquired a broad knowledge of the domain at stake and are capable of proposing a preliminary ontology, which can be enlarged and specified in subsequent steps.

In this particular research paper, adventure tourism was analysed from different approaches: the adventure activities provided and their features, the diverse textual genres, socio-economic factors and pragma-linguistic features[4]. Moreover, a set of domain experts were contacted and, thanks to their help, the following preliminary ontology was proposed (Figure 1):

[3] The author is grateful to the experts in physical and sport activities and active tourism that contributed to this research by providing their guidance and assistance in the development of this ontology.

[4] Due to space constraints, the results of this analysis are not presented here, but please see Durán-Muñoz (2014) for further information.

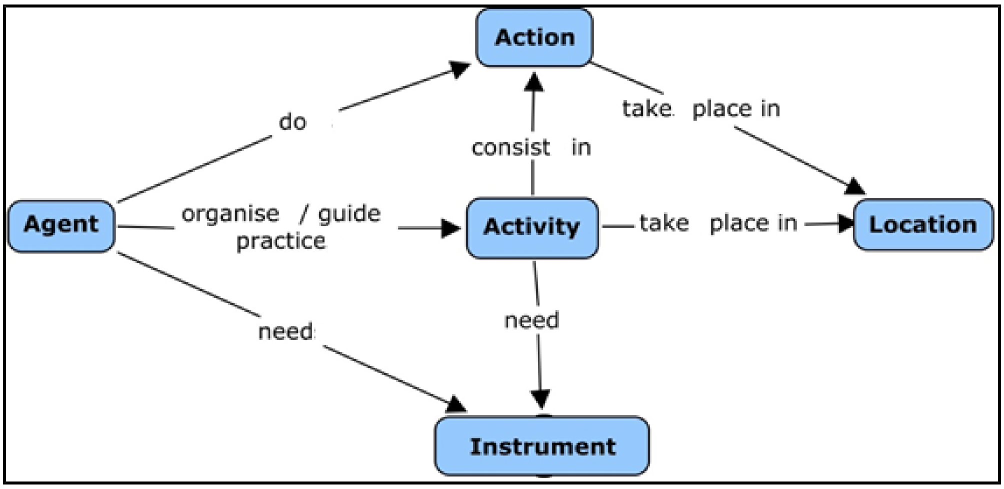


Figure 1
Preliminary ontology on adventure tourism.

This conceptual representation displays the prototypical situation in adventure tourism in a simplified way, indicating that all adventure activities (including canyoning) are organised in the same way. By means of this preliminary ontology, we can observe that, in order to understand and to define any concept within this frame, the activation of the other four categories that are part of this frame is also required. As such, an <agent>, i.e., a person who performs, guides or practises an activity is defined according to the activity and the actions which are performed along with the instruments employed for the activity; or for <activity>, it is necessary to explain the actions performed, the location in which it takes places and the instruments that are required. In short, the position of a concept in a particular context and with regard to other concepts is specified and, as a result, the ambiguity that may arise at language level disappears.

3.2. Corpus compilation

This step refers to the compilation of the specialised corpus, which provide conceptual and linguistic information to terminologists throughout the whole project. Consequently, the quality of this resource is of great importance to achieve successful results.

By way of an example, the semi-automatic compilation of a corpus about a particular adventure activity, canyoning, is provided in this subsection. This semi-automatic compilation was performed via *WebBootCat* tool (Baroni *et al*. 2006), a web-based automatic building tool integrated in the Sketch Engine online system[5] (Kilgariff *et al*., 2004). Even though the corpus was automatically compiled by the system, the result was carefully checked and manually revised in order to avoid irrelevant or inappropriate data that could bias the final analysis.

Following Baroni *et al*.'s protocol (2006), the compilation was divided into the following steps:

- Step 1. Seed words, i.e. keywords for the domain of interest, were selected. In this case, the seeds concerned the adventure activity canyoning (see Table 1) and were chosen after the analysis carried out in the previous step.

adventure activity	gorge
adventure	rope
outdoor activity	ravine
canyoning	waterfall
instructor	river
rappel	canyon
adventure	descend

Table 1
Most frequent terms in the adventure activity "canyoning", employed as seeds.

- Step 2. The seed words were used by the system to create "tuples" for the queries, i.e. they are randomly combined into different multi-word sets automatically by the programme (e.g. "canyoning river adventure").
- Step 3. The system generated a list of potentially relevant webpages, which could additionally be checked, deleted or confirmed according to the project needs, before building the corpus.
- Step 4. The selected webpages were automatically downloaded and put together in a single file (the corpus). They were also deduplicated and cleaned, and spam text or non-text was eliminated to obtain high-quality text material. At this stage, the corpus was available to be downloaded or browsed through Sketch Engine.

The same process can be repeated again and again to extend the corpus by means of the *Keywords/terms* option, which uses the keywords extracted from the compiled corpus as seed words to launch the search again.

[5] URL: https://www.sketchengine.eu/

The greatest advantage that this automatic corpus builder brings to corpus compilation is that very large corpora can be compiled quickly and effortlessly, compared to traditional time-consuming manual compilation. However, a thorough and careful manual revision of several aspects of the process is required to refine the searches and to guarantee successful results, which increases the total time of corpus building. Consequently, besides the clean-up processing that the tool carried out automatically, the proposed URLs in Step 3 were carefully revised and refined manually before the corpus building took place in order to discard any irrelevant URLs and to select only those that were appropriate for the project. As a matter of example, URLs discarded in this study were those of *Wikipedia*, *Amazon*, social networks (e.g. *Facebook*, *Pinterest*), *Youtube*, *Scribd*, *eBay*, etc.; those that were not originally written in English; and those that were not promotional texts, such as articles, blogs, etc., since the chosen textual genre for the corpus was promotional texts published by public and private companies and administration. After this manual revision, 30% of the URLs proposed by the system were discarded. Once revised and compiled, the English CANYON corpus, containing 925,422 words, was ready to be exploited in the next stages.

3.3. Term extraction

Once the compilation of the specialised corpus was complete, a semi-automatic term extraction was performed in *Sketch Engine* with the aim of extracting the most frequent units from the corpus. Figure 2 displays the most frequent single and multi-word units proposed by Sketch Engine, ordered according to their keyness.[6]

[6] The keyness is automatically calculated by Sketch Engine using the British National Corpus (BNC) as reference corpus.

Single-word		Score	F	RefF
rappel	W	1,487.04	2,370	10,127
canyoning	W	748.83	900	2,067
Canyoning	W	508.92	581	840
abseil	W	436.26	619	6,560
canyon	W	353.54	2,219	106,668
Lip	W	305.35	596	17,565
belay	W	287.18	472	11,217
Waterfall	W	272.32	499	15,112
Rappelling	W	268.20	458	12,543
waterfall	W	259.62	2,264	157,051
canyoneering	W	237.03	270	839
Sheer	W	236.89	318	5,028
Descent	W	222.48	328	7,752
Canyoneering	W	211.88	235	232

Multi-word		Score	F	RefF
subjective opinion	W	154.66	177	5
outdoor activity	W	85.46	104	13
outdoor adventure	W	78.62	91	7
activity instructor	W	74.43	81	0
zip line	W	68.99	75	0
white water	W	59.87	129	109
rappel device	W	59.02	64	0
full day	W	57.79	129	117
gorge walking	W	49.95	54	0
canyoning adventure	W	44.52	48	0
outdoor education	W	41.71	49	10
team building	W	41.45	56	28
outdoor recreation	W	41.41	47	6
adventure activity	W	41.10	45	2

Figure 2
Single and multi-word candidate terms, ranked by keyness.

The list of candidate terms obtained by means of the automatic keyword extraction tool gives an account of the most frequent units in this domain. However, it also requires a manual revision in order to detect the terms, which belong to this adventure activity, and to delete the units that do not, such as terms related to other fields that were also extracted by the tool (e.g. insurance-, accommodation-, and travelling-related terms), flora and fauna, or countries and nationalities.

After this manual revision, the terms extracted were thoroughly analysed and organised according to the five broad conceptual categories (Table 2), which coincide with the categories that were included in the preliminary ontology (Figure 1). These categories refer to the person (AGENT), the place (LOCATION), the action performed (ACTION) and the instrument employed (INSTRUMENT) in this adventure activity (ACTIVITY).

Agent	Activity	Action	Location	Instrument	
				Clothing	**Security**
Instructor	canyoning	trek	gorge	neoprene suit	harness
experienced instructor	rappelling	climb	mountain gorge	wetsuit	safety harness
professional instructor	canyoneering	walk	river gorge	glove	rope
qualified instructor	gorge walking	slide	valley gorge	mountain boot	safety rope
skilled instructor		swim	waterfall	wetsuit boot	single rope
expert staff		rappel	(natural) pool	rock shoe	double rope
canyoner		abseil	natural park	waterproof trouser	eight
		descent	national park		helmet
			river		safety helmet
			puddle		karabiner / carabiner
			ravine		rappel device
					zip line

Table 2
Categorisation of selected terms after term extraction.

3.4. Building corpus-based domain ontologies

The preliminary frame-based ontology created during the first step of this ontoterminography methodology can now be completed with the terms extracted from the CANYON corpus. Hence, Figure 3 shows the combination of the top-down (step 1) and bottom-up (step 2) of this methodology.

By carrying out a further corpus-based bottom-up study, consisting of searching for terms in context with a KWIC tool (such as *AntConc*)[7] and studying their concordances and co-text, the recurrent linguistic patterns of the conceptual relations were identified (Table 3). This method was also enriched with top-down *in vitro* analysis, i.e. study of definitions included in terminological resources about canyoning[8] and domain expert consultation.

[7] *AntCon* is a freeware concordancer developed by Laurence Anthony. URL: http://www.antlab.sci.waseda.ac.jp/software.html
[8] There are very few terminological resources about canyoning, but some online glossaries including terminology related to this activity can be found. For example, *Dictionary of terms used in canyoning* (https://canyonmag.net/es/technical/essentials/dictionary/), *Glossary of canyoneering terms* (http://dyeclan.com/outdoors101/canyoneering101/?page=glossary-of-

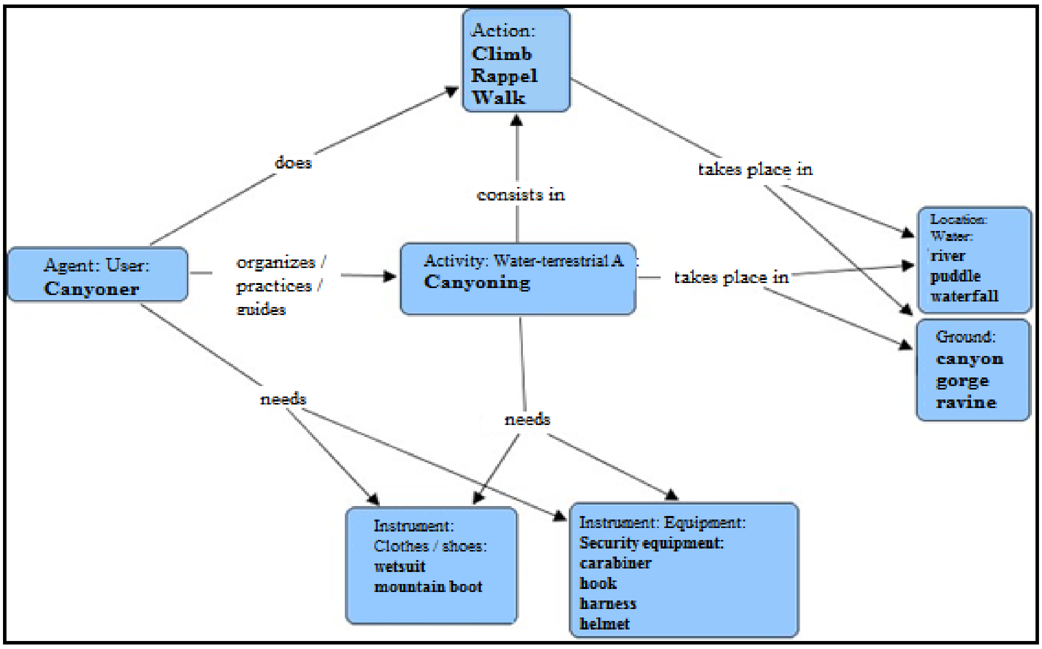


Figure 3
Preliminary frame-based ontology with canyoning-related terms (in bold).

With the acquisition of this further conceptual knowledge from the corpus, the preliminary conceptual representation was enlarged and became an extended corpus-based ontology. During this process, some subcategories were added (see Annex 1):

- In <agent>, three different agents were identified: 1. <organiser>, the person in charge of organising the adventure activity; 2. <technical staff>, dedicated to lead, coach or guide the activities, and 3. <user>, which corresponds to the person practicing the activity.
- In <instrument>, 1. <clothes/shoes> and 2. <security equipment> were identified, which were further subcategorised.
- <location> was also divided in different types of location (<water-air>, <air>, <water-ground>, <water>, <ground>).
- Category <activity> also was subdivided into <underground activity>, <water activity>, etc.

As observed, the extended corpus-based ontologies that result from the application of the ontoterminography methodology helps terminographers

canyoneering-terms) or *Canyoneering Glossary* (http://www.outdoorstogether.com/canyoneering_glossary.htm).

to organise the specialised knowledge in a clear and coherent way and becomes crucial in the subsequent steps of the methodology.

CONCEPTUAL RELATIONS	LINGUISTIC PATTERNS	CONCEPTUAL RELATIONS	LINGUISTIC PATTERNS
IS_A	- is a* - is a kind of - is a variant of - is a sort of - is a type of - is/are called - is an activity - such as - for example - known as - similar to - based on - named - a variety of	PART_OF	- contain* - composed of - compris* - is / are comprised of - a range of - consist* of - completed by - complemented by - classified - include* - including - is / are included
CONSIST_IN	- involv* - requir* - based on - with the aim of	TAKE_PLACE	- found at - found in - found throughout - in contact with - in the heart of - where - known in - located on - located in - situated in - situated on - tak* place - practi* on - practi* in - climb* up
NEED	- is / are required - required for - use* - using - the use of - mak* use of - utilis* - practis* with - is / are necessary - need* for - by means of - equipped with - be of help - with the help of - launch* by	ORGANISE GUIDE	- taught by - teach* - offer* - organis* - guid* - help* - lead* - made by - with the support of - the service* of - is / are responsible for - in the safe hands of - recommend* - advis*
		PRACTISE	- practi*ed by - done by

Table 3
Linguistic patterns of conceptual relations.

4. Final remarks

This paper applies the ontoterminography methodology to illustrate the different steps that are required to build a corpus-based ontology in a specialised domain, in this case the adventure activity of canyoning. It describes the different stages in which this middle-out protocol is divided and focuses on the first three, namely: 1. Design of the project and analysis of the domain, 2. Corpus compilation, 3. Term extraction and 4. Building of a corpus-based ontology.

As it has been proved, corpus-based ontologies in specialised domains provide users with a clear and organised representation of the specialised knowledge of a domain, including the main concepts and the conceptual relations among them, and are a convenient starting point to produce specialised resources. Corpora, for its part, are crucial in this methodology, both in the conceptual and linguistic dimension of any terminological project, since they are relevant information sources that contribute to the enlargement of conceptual representation as well as to the identification of terms, contexts, synonyms, equivalents, etc. This makes them indispensable for any terminological project nowadays.

With the aim of applying the methodology presented in this paper, the author is currently working on *DicoAdventure*, an online multilingual terminological database about adventure tourism which will contain terminology concerning the most common adventure activities, including canyoning, in English and Spanish. This project is the result of a collaboration with the OLST (*Observatoire de linguistique Sens-Texte*) at the University of Montreal.

Finally, it is worth mentioning that specialised knowledge is continuously evolving and changing over time, due to the emergence of new concepts, the change in meaning of previous concepts or the deletion of some others. Therefore, terminographers should not see any knowledge representation as stable and finite, but as a dynamic resource that may require adjustments in time.

Bionote: Lecturer (with tenure track) at the Department of English and German Philologies since 2016. BA in Translation and Interpreting from the University of Málaga and BA in English Studies from UNED. She obtained her PhD in June 2011, which was awarded with the best PhD dissertation award in the IV Biennial Iberian Association of Translation and Interpreting Studies [AIETI] and the Special doctorate award. Researcher in European, national, and regional R&D projects, and invited researcher at the Research Institute in Information and Language Processing (RIILP) of the University of Wolverhampton (UK), Erasmushogeschool Brussels (Belgium) and University of Montreal (Canada), among others. She has published in well-known national and international relevant conferences and journals. Some of her publications have received

awards for the quality of her research. Her main research lines are corpus linguistics, terminology, lexico-semantics, linguistic technologies and ICTs for foreign language teaching.

Author’s address: iduran@uco.es

Acknowledgements: This research was partially carried out within the framework of the project *VIP: sistema integrado Voz-texto para IntérPretes* (ref. n. FFI2016-75831, 2017-2020) and the thematic network *TRAJUTEC: TRAducción JUrídica y TECnología* (University of Málaga).

References

Atkins S. and Rundell M. 2008, *The Oxford Guide to Practical Lexicography*, Oxford University Press, Oxford.

Baroni M., Kilgarriff A., Pomikálek J. and Rychlý P. 2006, *WebBootCaT: a web tool for instant corpora*, in *Proceedings of Euralex 2006*, Torino, Italy.

Cabré Castellví M.T. 1999, *Hacia una teoría comunicativa de la terminología: aspectos metodológicos*, in Cabré M.T. (ed.), *La Terminología: Representación y Comunicación. Elementos para una teoría de base comunicativa y otros artículos*, IULA, Universidad Pompeu Fabra, Barcelona, pp. 129-150.

Dash N.S. 2000, *Corpus Linguistics: An Introduction*, Pearson Longman, New Delhi.

Durán-Muñoz I. 2012, *La ontoterminografía aplicada a la traducción. Propuesta metodológica para la elaboración de recursos terminológicos dirigidos a traductores*, Peter Lang, Berlin.

Durán-Muñoz I. 2014, *Aspectos pragmático-lingüísticos del discurso del turismo de aventura: estudio de un caso*, in "Revista Normas" 4, pp. 49-69.

Faber P., Márquez Linares C., and Vega Expósito M. 2005, *Framing Terminology: A Process-Oriented Approach*, in "Meta" 50 [4], pp. 1-10.

Fillmore C. 1968, *The Case for Case*, in Bach E. and Harms R. (eds.), *Universals in Linguistic Theory*, Holt, Rinehart and Winston, New York, pp. 1-88.

Gruber T.R. 1993, *A Translation Approach to Portable Ontologies*, in "Knowledge Acquisition" 5 [2], pp. 199-220.

ISO (International Organization for Standarization) 2000, *ISO 704: 2000. Terminology Work. Principles and methods*, ISO, Genève.

ISO (International Organization for Standarization) 2000, *ISO 1087-1:2000. Terminology work – Vocabulary – Part 1: Theory and application*, ISO, Genève.

Leech G. 1992 *Corpora and theories of linguistic performance*, in Svartvik J. (ed.), *Directions in corpus linguistics. Proceedings of Nobel Symposium 82*, Mouton de Gruyter, Berlin/New York, pp. 105-122.

L'Homme M.C. 2008, *Le DiCoInfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés*, in "Traduire" 217, pp. 78-103.

Moreno Ortiz A. 2004, *Representación de la información terminológica en Ontoterm®: Un sistema gestor de bases de datos terminológicas basado en el conocimiento*, in Faber Benítez P. and Jiménez Hurtado C. (eds.), *Investigar en terminología*, Comares, Granada, pp. 25-70.

Roche C. 2007, *Saying is not modelling*, in *Proceedings of the 4th International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2007)*, Funchal, Portugal, pp 47-56.

Roche C., Calberg-Challot M., Damas L. and Rouard P. 2009, *Ontoterminology: A new paradigm for terminology*, in *Proceedings of International Conference on Knowledge Engineering and Ontology Development, KEOD 2009*, Madeira.

Temmerman R. 2000, *Towards New Ways of Terminology Description. The sociocognitive approach*, John Benjamins, Amsterdam.

Temmerman R. and Kerremans K. 2003, *Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description*, in Hajicov E., Kotešovcov A. and Mírovský, J. (eds.), *Proceedings of the 17th International Congress of Linguists (CIL17)*, Matfyzpress, MFF UK (CD-ROM), Prague.

ANNEX 1. Extended semantic frame (or domain ontology)