

Lingue & Linguaggi

65/2024

Special Issue

**Transformaciones de los enfoques de
investigación en el discurso digital**

**Transformationen linguistischer Ansätze
im digitalen Diskurs**

**Transformations des approches
linguistiques des discours numériques**

Edited by

Claudia Cagninelli, Daniela Capra

Ulrike A. Kaunzner, Chiara Preite



UNIVERSITÀ
DEL SALENTO

2024

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 65/2024

COMITATO DI CONSULENZA SCIENTIFICA DELLA RIVISTA

Orietta Abbati, Università di Torino
Jörn C. Albrecht, Rupprecht-Karls-Universität Heidelberg
Pedro Álvarez de Miranda, Real Academia Española
Carmen Argondizzo, Università della Calabria
Sara Augusto, Universidade de Coimbra
Gabriele Azzaro, Università di Bologna
Marcos Bagno, Universidade de Brasília
Carla Barbosa Moreira, CEFET- MG - Brasile, Brazil
Jean-Philippe Barnabé, Université de Picardie (Amiens, Francia), France
Simona Bertacco, University of Louisville, USA
Silvia Betti, Alma Mater-Università di Bologna e ANLE (Academia Norteamericana de la Lengua Esp.), Italy
Francesco Bianco, Univerzita Palackého v Olomouci
Alessandro Bitonti, Masaryk University, Czech Republic
Jean-Claude Blachère, Université Paul Valéry, Montpellier III
Maria Bortoluzzi, Università degli Studi di Udine
Chiara Bolognese, Universidad Autónoma de Barcelona
Lorella Bosco, Università degli Studi di Bari
Maria Luiza Braga, Universidade Federal do Rio de Janeiro
Cristiano Broccias, Università degli Studi di Genova
Silvia Bruti, Università degli Studi di Pisa
Sandra Campagna, Università degli Studi di Torino
Catherine Camugli Gallardo, Université Paris Ouest - Nanterre
Xelo Candel Vila, Universitat de València
Martha Canfield, Università degli Studi di Firenze
Gloria Cappelli, Università di Pisa, Italy
Manuel Carrera Díaz, Universidad de Sevilla
Alessandro Cassol, Università degli Studi di Milano
Gabriella Catalano, Università di Roma "Tor Vergata"
Paola Catenaccio, Università degli Studi di Milano
Marco Cipolloni, Università di Modena e Reggio Emilia
Carmen Concilio, Università degli Studi di Torino
Alessandro Costantini, Università degli Studi di Venezia
Pier Luigi Crovetto, Università degli Studi di Genova
Maria Debora De Fazio, Università degli Studi della Basilicata
Giorgio de Marchis, Università Roma Tre
Maria del Valle Ojeda Calvo, Università degli Studi di Venezia "Ca' Foscari"
Arthur DeFrance, Università di Waseda, Tokyo
Jean-Christophe Delmeule, Université Charles De Gaulle, Lille 3
Gabriella Di Martino, Università degli Studi di Napoli
Marina Dossena, Università degli Studi di Bergamo, Italy
Jean-François Durand, Université Paul Valéry, Montpellier III
Claus Ehrhardt, Università degli Studi di Urbino, Italy
Roberta Facchinetti, Università degli Studi di Verona
Federica Ferrari, Università degli Studi di Bologna
Teresa Ferrer Valls, Universitat de València
Luisanna Fodde, Università degli Studi di Cagliari
Giuliana Elena Garzone, Libera Università IULM di Lingue e Comunicazione
Sara Gesuato, Università degli Studi di Padova
Barbara Gori, Università degli Studi di Padova
Enrico Grazi, Università degli Studi Roma 3
Dorothee Heller, Università degli Studi di Bergamo
Laeticia Jensen Eble, Universidade de Brasília
Mersini Karagevrekci, University of Macedonia
Jean René Klein, Université catholique de Louvain

Emil Lafe, Centro di Studi Albanologici, Tirana
Elena Landone, Università di Sassari
Anna Maria Laserra, Università degli Studi di Salerno
Lucilla Lopriore, Università degli Studi Roma 3, Italy
Monica Lupetti, Università di Pisa
Stefania Maci, Università degli Studi di Bergamo
Aldo Antonio Magagnino, Professional literary translator, Italy
Francisco Martín, Università degli Studi di Torino
Antonio Manieri, Università di Napoli "L'Orientale"
Federico Masini, "Sapienza" Università di Roma
Daniela Mauri, Università degli Studi di Milano
Selena Millares, Universidad Autónoma de Madrid
Sandro M. Moraldo, Università di Bologna
Rafael Morales Barba, Universidad Autónoma de Madrid, Spain
Marina Morbiducci, Università di Roma Sapienza
Mara Morelli, Università degli Studi di Genova
Francesco Morleo, Italy
Martina Nied, Università di Roma Tre
Liana Nissim, Università degli Studi di Milano
Vincenzo Orioles, Università degli Studi di Udine
Elisa Perego, Università degli Studi di Trieste, Italy
Alain Peyraube, Ecole des Hautes Etudes en Sciences Sociales (EHESS, Paris) and Emeritus director of research at the Centre National de la Recherche Scientifique (CNRS, Paris, France)
Barbara Sasse, Università degli Studi di Bari "Aldo Moro"
Francesco Saverio Perillo, Università degli Studi di Bari
Ronny F. Schulz, Christian-Albrechts-Universität su Kiel
Elena Pessini, Università degli Studi di Parma
Diane Ponterotto, University of Rome Tor Vergata
Franca poppi, Università di Modena e Reggio Emilia
Chiara Preite, Univ. di Modena e Reggio Emilia
Virginia Pulcini, Università di Torino
Lupe Romero Ramos, Universidad Autónoma de Barcelona
Alessandra Riccardi, Università di Trieste
Silvia Riva, Università degli Studi di Milano
Federica Rocco, Università degli Studi di Udine
José-Carlos Rovira Soler, Universidad de Alicante
Mette Rudvin, Università di Bologna, Italy
Vincenzo Russo, Università di Milano
Rita Salvi, Università di Roma "La Sapienza"
Antonio Sánchez Jiménez, Université de Neuchâtel
Julián Sauquillo Gómez, Universidad Autónoma de Madrid
Michael Schreiber, Johannes-Gutenberg-Universität Mainz
Marcello Soffritti, Università degli Studi di Bologna
Elena Spandri, Università degli Studi di Siena
Valeria Tocco, Università di Pisa
Ilda Tomas, Università di Granada, Spain
Georgina Torello, Universidad de la República (Montevideo)
Nicoletta Vasta, Università di Udine
Germán Vega García-Luengos, Universidad de Valladolid
Ivan Verc, Università degli Studi di Trieste
Graciela Villanueva, Université de Paris Est Créteil Val de Marne
Itala Vivan, Università degli Studi di Milano
Bryan Weston Wyly, Università della Val D'Aosta
Raúl Zamorano Farias, Universidad Nacional Autónoma de México

DIRETTORE RESPONSABILE: Maria Grazia Guido, Università del Salento

DIRETTORE DEL DIPARTIMENTO DI STUDI UMANISTICI E DIRETTORE SCIENTIFICO: Maria Grazia Guido, Università del Salento

JOURNAL MANAGER: Pietro Luigi Iaia, Università del Salento

COMITATO EDITORIALE: Marcello Aprile, Yasemin Bayyurt, Miguel Bernal-Merino, Francesca Bianchi, Gualtiero Boaglio, Victoria Bogushevskaya, Thomas Christiansen, Alessia Cogo, Rosita D'Amora, Giulia D'Andrea, Antonella De Laurentiis, Maria Luisa De Rinaldis, Gian Luigi De Rosa, Martin Dewey, Jorge Díaz Cintas, Giulia Andreina Disanto, Maria Renata Dolce, Monica Genesin, Maria Teresa Giampaolo, Barbara Gili Fivela, Mirko Grimaldi, Maria Grazia Guido, Gerhard Hempel, Pietro Luigi Iaia, Marcella Leopizzi, Elena Manca, Simona Mercantini, Maria Chiara Migliore, Fabio Moliterni, Antonio Montinaro, Rocco Luigi Nichil, Gloria Politi, Luciano Ponzio, Mariarosaria Provenzano, Irene Romera Pintor, Amalia Rodríguez Somolinos, Alessandra Rollo, Andrea Scardicchio, Virginia Sciutto, Diego Simini, Carmelo Spadola, Anne Theissen, Elena I. Trofimova, Natasha Tsantila.

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.

© 2024 University of Salento - Coordinamento SIBA

<http://siba.unisalento.it>

ISSN 2239-0367

eISSN 2239-0359 (electronic version)

<http://siba-ese.unisalento.it>



Indice

- 7 CLAUDIA CAGNINELLI, DANIELA CAPRA, ULRIKE A. KAUNZNER, CHIARA PREITE, *Transformaciones de los enfoques de investigación en el discurso digital; Transformationen linguistischer Ansätze im digitalen Diskurs; Transformations des approches linguistiques des discours numériques*

PARTE – TEIL – PARTIE I

TRANSFORMACIONES DIGITALES EN LA ADQUISICIÓN LINGÜÍSTICA Y TERMINOLÓGICA

DIGITALE TRANSFORMATIONEN IM SPRACH- UND TERMINOLOGIEERWERB

TRANSFORMATIONS NUMERIQUES DANS L'ACQUISITION LINGUISTIQUE ET TERMINOLOGIQUE

- 39 INES BOSE, STEPHANIE KURTENBACH, *„aber das WAR AUCH n bisschen SCHÖN bei dem traum“. Exemplarische Studie zum Sprechen über inneres Erleben in Gesprächskreisen des Kindergartens*
- 67 ULRIKE A. KAUNZNER, LISA BARDECK, *Die Datenbank „Studentische Referate in der Fremdsprache Deutsch“ und ihre Anwendung in einer Pilotstudie zur rhetorischen Sprechwirkung*
- 99 CHRISTIAN GEGNER, *Analyse der stimmlichen Leistungsfähigkeit bei Lehramtsstudierenden. Ein Vergleich zweier digital gestützter multiparametrischer Verfahren*
- 127 CARLOS ROCHA OCHOA, *Los glosarios financieros electrónicos en español y portugués como instrumentos de educación financiera en Latinoamérica*

PARTE – TEIL – PARTIE II

TRANSFORMACIONES DIGITALES EN LAS FORMAS LINGÜÍSTICAS Y DISCURSIVAS DEL PATRIMONIO CULTURAL

DIGITALE TRANSFORMATIONEN DER SPRACHLICHEN UND DISKURSIVEN FORMEN DES KULTURELLEN ERBES

TRANSFORMATIONS NUMERIQUES DANS LES FORMES LINGUISTIQUES ET DISCURSIVES DU PATRIMOINE CULTUREL

- 149 ELENA CARPI, *Dicciocho. An 18th Century Lexicographic Portal*
- 163 JOSÉ FRANCISCO MEDINA MONTERO, *Algunas calas en los diez primeros capítulos de la primera parte del Don Chisciotte de Franciosini: “toscanismos”, religión y germanía*
- 183 SIMONE TEPEDINO, *Discurso museístico e identidad: una comparación*

PARTE – TEIL – PARTIE III

TRANSFORMACIONES DIGITALES EN EL ANÁLISIS LÉXICO, FRASEOLÓGICO Y TERMINOLÓGICO

DIGITALE TRANSFORMATIONEN IN DER LEXIKALISCHEN, PHRASEOLOGISCHEN UND TERMINOLOGISCHEN ANALYSE

TRANSFORMATIONS NUMERIQUES DANS L'ANALYSE LEXICALE, PHRASEOLOGIQUE ET TERMINOLOGIQUE

- 211 ANDRÉS ORTEGA GARRIDO, *Presencia de léxico coloquial en las entrevistas de Jordi Évole: estudio de corpus*
- 233 JENS FLEISCHHAUER, *Eine datengesteuerte Untersuchung deutscher Funktionsverbgefüge. Artikelvarianz in stehen-Funktionsverbgefügen*

- 267 PAOLO FRASSI, *Entre terminologie, terminographie et humanités numériques : le projet DIACOM-fr*
- 285 CHIARA PREITE, *Les dénominations des produits typiques de l'Émilie-Romagne. Diffusion multilingue d'une terminologie protégée*
- 303 MICHELA TONTI, *Quelle souplesse pour le signe linguistique du Nom de Marque ? Étude de sa variation phraséologique au prisme de la créativité du discours numérique au quotidien*
- 327 DANIELA CAPRA, *Usos y fraseología de la palabra "pelo" en castellano. Un estudio de corpus hasta el siglo XVII*

PARTE – TEIL – PARTIE IV

TRANSFORMACIONES DIGITALES EN EL ANÁLISIS DEL DISCURSO

DIGITALE TRANSFORMATIONEN IN DER DISKURSANALYSE

TRANSFORMATIONS NUMERIQUES DANS L'ANALYSE DU DISCOURS

- 361 JULIEN LONGHI, *Explorer les mots du politique dans la transformation numérique. Analyser le lexique politique dans des contextes et selon des ressources en évolution*
- 385 CLAUDIA CAGNINELLI, *Discours numériques et approches outillées : quelques réflexions sur les apports des métadonnées*
- 413 MARCELLA PALLADINO, *Webbasierte Tools für die Transkription und Analyse von Reden. Hilfreiche Instrumentarien für die (Polito)Linguistik*
- 439 ISABELLA FERRON, *Digitalisierung der politischen Kommunikation. Eine politolinguistische Analyse der Twitter/X-Kommunikation ausgewählter AfD-Politiker*
- 463 GLORIA ZANELLA, *Les métaphores no-vax sur Twitter. Une étude comparative français-italien*

INTRODUCCIÓN – EINFÜHRUNG – INTRODUCTION

CLAUDIA CAGNINELLI¹, DANIELA CAPRA²,
ULRIKE A. KAUNZNER³, CHIARA PREITE³

¹UNIVERSITÀ DEGLI STUDI DI MILANO

²UNIVERSITÀ DEGLI STUDI DI GENOVA

³UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

TRANSFORMACIONES DE LOS ENFOQUES DE INVESTIGACIÓN EN EL DISCURSO DIGITAL

La revolución digital constituye uno de los eventos recientes que más han influido en las transformaciones lingüísticas, debido a la evolución y los cambios de las prácticas discursivas mismas en ese contexto. Los estudios sobre el discurso digital, cada vez más numerosos desde finales del siglo pasado, han mostrado el impacto del uso de las nuevas tecnologías en todos los niveles de las producciones verbales y aún más en las multimodales. Las transformaciones introducidas por lo digital se multiplican en las prácticas lingüísticas ligadas a los nuevos instrumentos de comunicación e interacción en línea – como los forums (Mourlhon-Dallies *et al.* 2004), los blogs y sobre todo las redes sociales (Paveau 2017; Zappavigna 2018) –, y asimismo al uso de recursos y formas de comunicación específicas, como, por poner un par de ejemplos, los hashtags (Zappavigna 2015; Mercier 2018) y los memes (Davison 2012; Simon, Wager 2023). Además, los nuevos instrumentos digitales han contribuido a dar espacio y voz a una cantidad cada vez más grande de usuarios nacionales e internacionales, institucionales, económicos, sociales, pero también – y sobre todo – a la opinión pública, gracias a la introducción del web participativo.

Consiguientemente, las transformaciones lingüísticas y discursivas vinculadas al recurso a las tecnologías digitales han causado repercusiones en las perspectivas teóricas y metodológicas utilizables en el estudio de los discursos digitales, modificándolas e innovándolas a partir de la concepción misma de textualidad (Mayeur, Paveau 2020), del discurso y de los géneros discursivos (Paveau 2017). El lenguaje verbal, en efecto, no es sino uno de los componentes del discurso digital, cuya naturaleza multimodal y multimedial ha ampliado el campo de indagación de los estudios lingüísticos y discursivos, que ahora se dirigen con una frecuencia cada vez mayor hacia el estudio de la construcción plurisemiótica y multimodal de estas prácticas (Kress, van Leeuwen 2001; Paveau 2017).

Desde hace varios años la investigación lingüística ha resaltado ampliamente la vitalidad y la complejidad de las producciones discursivas, subrayando los diferentes elementos lingüísticos y extralingüísticos que participan en sus constantes y múltiples transformaciones. Debido a la variedad y la pluralidad de los factores lingüísticos y extralingüísticos que determinan específicas complejidades, en el análisis de los discursos digitales se revela aún más necesario un acercamiento contextualizado al estudio de las producciones verbales. Las transformaciones de los discursos digitales presentan además rastros de la evolución social y cultural de las correspondientes prácticas comunicativas, que destacan tanto en el nivel léxico-semántico, como fraseológico y, más en general, textual y discursivo.

Si las transformaciones lingüísticas y discursivas del ámbito digital han descubierto la necesidad de adaptar los enfoques teórico-metodológicos existentes, estos últimos a su vez han beneficiado de las recientes innovaciones tecnológicas y digitales. El recurso a herramientas informáticas para el análisis lingüístico y discursivo ha contribuido al desarrollo de nuevas perspectivas de investigación y de análisis, revolucionando los enfoques lingüísticos dirigidos al estudio del lenguaje. Nuevos ámbitos y metodologías han hecho de los recursos informáticos un aspecto fundamental del estudio lingüístico, como por ejemplo en el caso de los estudios de corpus (Habert *et al.* 1997; Biber *et al.* 1998; Williams 2005; McEnery, Hardie 2012), y de las aportaciones, en ámbito francófono, de la textometría (Mayaffre *et al.* 2019) y de la estadística textual en el análisis de datos (Lebart *et al.* 2019) y en ámbito hispánico de la estilometría (Cuéllar, Vega García-Luengos 2017-2024) y de la creación de portales interrogables (véase Carpi en este volumen). La confección de corpus digitales e interrogables a través de programas informáticos es central hoy en día en muchos estudios lingüísticos y discursivos, como demuestran varios artículos recogidos en estas páginas. También en el ámbito lingüístico se está incrementando la implementación de recursos digitales auxiliados por las potencialidades de la inteligencia artificial, que se integra en el análisis lingüístico y discursivo (Mayaffre, Vanni 2021; Longhi y Palladino en este volumen).

Este número monográfico recoge dieciocho contribuciones que tienen como objeto de estudio tres lenguas: alemán, español y francés. La diversidad lingüística que caracteriza el volumen pretende promover una reflexión translingüística y transcultural a través de la organización de las distintas contribuciones de acuerdo con cuatro ejes temáticos principales:

1. transformaciones digitales en la adquisición lingüística y terminológica;
2. transformaciones digitales en las formas lingüísticas y discursivas del patrimonio cultural;
3. transformaciones digitales en el análisis léxico, fraseológico y

terminológico;

4. transformaciones digitales en el análisis del discurso.

Los artículos se han agrupado según el tema, abarcado desde diferentes perspectivas, sin tener en cuenta el idioma en el cual están escritos. Once de ellos constituyen unos *selected papers* ligados al congreso organizado por el Centro Interuniversitario de Investigación CLAVIER 2021 en colaboración con los grupos de investigación *Lexi-Term* y *CAP (Professional and Academic Communication)*, titulado *Exploring Words in the Digital Transformation. Tools and Approaches for the Study of Lexis and Phraseology in Evolving Discourse Domains*, que tuvo lugar en la Universidad de Módena y Reggio Emilia en los días 18 y 19 de noviembre de 2021. El presente número, por tanto, se pone en diálogo directo con los estudios anglófonos de dicho congreso, que se recogieron en el número 58 de *Lingue e Linguaggi* (Bondi, Cacchiani 2023) y en el número 27 de *Status Quaestionis* (Bondi, Cacchiani 2024).

Los artículos aquí reunidos que forman parte de los que se presentaron en el Congreso Clavier 2021 se centran en el impacto de las transformaciones digitales según perspectivas distintas pero profundamente interconexas, y muestran la interrelación de fenómenos léxico-fraseológicos y socio-discursivos. Siete artículos más, seleccionados sucesivamente, pretenden ofrecer una contribución a la reflexión alrededor de los ejes temáticos arriba mencionados, ampliando ulteriormente los objetos de estudio analizados en el núcleo original.

En su conjunto, el presente volumen se pone pues el objetivo de analizar las profundas transformaciones introducidas por las tecnologías digitales en las prácticas lingüísticas y discursivas tanto en ámbitos de la lengua general, como en ámbitos especializados; pretende asimismo estudiar las perspectivas teóricas y metodológicas adoptadas para el análisis tanto de la comunicación que se desarrolla en un ambiente digital, como de los datos digitalizados y de los explorados a través de instrumentos digitales. Por un lado, los artículos abordan el profundo cambio *in fieri* en los medios digitales en evolución, con nuevas formas de participación y contribución; por el otro, dirigen su atención al papel y al aporte de las herramientas digitales en el estudio del léxico y de la fraseología, o en el del desarrollo de la lexicografía, la lexicología, la semántica léxica y el análisis del discurso. En estos casos, los trabajos se focalizan ya en la digitalización de obras antiguas, ya en la construcción y la exploración de conjuntos de textos o corpus recientes, cada vez más enjundiosos y extendidos.

En el primer eje de investigación, dedicado a las *Transformaciones digitales en la adquisición lingüística y terminológica*, los estudios se centran en el impacto de las transformaciones digitales en el desarrollo de las competencias lingüístico-discursivas y terminológicas en cuatro contextos de aprendizaje

diferentes: el ámbito pre-escolar (Bose/Kurtenbach), el universitario (Kaunzner/Bardeck), el de la formación de profesores (Gegner) y finalmente de la educación de los ciudadanos (Rocha Ochoa). En esta primera parte, se muestra la aportación de herramientas digitales y de metodologías para el análisis del aprendizaje, y el uso de recursos digitales como herramienta para aprender.

El artículo de **Ines Bose** y **Stephanie Kurtenbach**, „*aber das WAR AUCH n bisschen SCHÖN bei dem traum*“. *Exemplarische Studie zum Sprechen über inneres Erleben in Gesprächskreisen des Kindergartens*, nace de una indagación en los círculos de conversación de una escuela materna alemana, cuyo objetivo es la promoción de la participación social y el desarrollo lingüístico. La investigación se basa en un corpus de vídeos, que documenta las interacciones entre los niños. Las autoras presentan un estudio de caso sobre las estrategias de resolución de problemas entre niños de seis años, y se focalizan sobre la capacidad infantil de expresar la experiencia psicológica interna (emociones, volición, cognición), crucial para la comunicación interpersonal. Siguiendo los postulados teóricos, se analizan como ejemplo los datos de un colectivo de conversación, mostrando el empleo de medios lingüísticos para hablar de emociones y estados interiores, medios adquiridos a través de una conversación empática auxiliada por profesionales.

En su artículo, titulado *Die Datenbank „Studentische Referate in der Fremdsprache Deutsch“ und ihre Anwendung in einer Pilotstudie zur rhetorischen Sprechwirkung*, **Ulrike A. Kaunzner** y **Lisa Bardeck** presentan y discuten la elaboración de una base de datos formada por grabaciones audio y video de presentaciones hechas por los estudiantes de alemán como lengua extranjera, y reflexionan sobre las diferentes etapas de la transcripción con la herramienta OCTRA, subrayando las potencialidades y las limitaciones de su aplicación. Un estudio longitudinal de un sub-corpus de videograbaciones a distancia recogidas a lo largo de dos años revela cambios interesantes en las habilidades verbales de los estudiantes y compara autoevaluaciones con evaluaciones externas. El análisis, que se centra en las diferencias entre presentaciones presenciales y a distancia, se concentra además en aspectos como la impresión general, la claridad, la naturalidad, la competencia percibida y la sustancia de la expresión, focalizándose en particular en algunos rasgos fonéticos. Las autoras dirigen asimismo su atención a posibles perspectivas de investigación y didácticas.

En *Analyse der stimmlichen Leistungsfähigkeit bei Lehramtsstudierenden. Ein Vergleich zweier digital gestützter multiparametrischer Verfahren*, **Christian Gegner** centra su atención en la voz de los docentes, una categoría profesional expuesta a esfuerzos vocales elevados. Sin embargo, muchos profesionales no están adecuadamente preparados para enfrentarse con esta tarea, lo que lleva a una significativa

prevalencia de problemas vocales. El artículo propone un análisis vocal digital y coteja dos métodos multi-paramétricos, con un muestreo de 200 estudiantes de la Universidad de Ratisbona (Alemania). La conclusión explora las posibilidades de desarrollo de una aplicación para la higiene y la prevención vocal para los profesionales de la enseñanza.

El trabajo de **Carlos Rocha Ochoa**, *Los glosarios financieros electrónicos en español y portugués como instrumentos de educación financiera en Latinoamérica*, concluye la primera esfera de investigación del volumen, favoreciendo además la transición hacia la segunda parte, que se centra más específicamente en el léxico. A través del análisis comparativo de obras lexicográficas digitales, Carlos Rocha Ochoa indaga su papel en la promoción de la educación y la inclusión financieras en América Latina. Contrariamente a los trabajos anteriormente mencionados, la atención se dirige aquí a la alfabetización del ciudadano común en un terreno especializado que, sin embargo, afecta a la sociedad en su conjunto. Los cinco glosarios financieros electrónicos estudiados por Carlos Rocha Ochoa se analizan en una perspectiva funcional, tomando en consideración, además de la estructura, las necesidades de los usuarios de los servicios financieros, la accesibilidad y la metodología de elaboración, incluidos los elementos metalexográficos como los objetivos y las explicaciones de los autores. Además de mostrar la disparidad que caracteriza los glosarios analizados por lo que se refiere a la accesibilidad y a la relevancia de los términos seleccionados, el análisis de los cinco glosarios constituye, según Carlos Rocha Ochoa, un punto de partida para avanzar en este terreno, con el objetivo de mejorar la comprensibilidad de las definiciones.

La lexicografía tiene un papel determinante también en la creación y transformación del patrimonio cultural del que el idioma es testigo. El rol de las palabras como canal de acercamiento y documento de la renovación de las construcciones y representaciones culturales e identitarias es el tema de la segunda parte del volumen, *Transformaciones digitales en las formas lingüísticas y discursivas del patrimonio cultural*, que recoge tres artículos.

La segunda parte del volumen se abre con un artículo dedicado al análisis lexicográfico, y más precisamente a la presentación de una herramienta diseñada para los diccionarios españoles del siglo XVIII, el portal lexicográfico *Dicciocho*, creado por el *Grupo de investigación Dicionarística del siglo XVIII* del que forma parte **Elena Carpi**. En su artículo, *Dicciocho. An 18th Century Lexicographic Portal*, Elena Carpi subraya el interés del portal como instrumento de divulgación y de acceso para la comunidad científica al lecionario de la segunda edición del *Diccionario de Autoridades*, considerado durante largo tiempo perdido y solo en épocas recientes encontrado en los archivos de la Real Academia Española. La autora hace además constar el continuo ampliamento del portal, gracias a la integración de materiales relativos a la primera edición de dicho diccionario

(1726-39) y al *Diccionario castellano con voces de Ciencias y Artes* de Terreros y Pando. A través de voces ligadas al ámbito artístico, en particular arquitectura y pintura, la autora ilustra, en este artículo, el uso del portal *Dicciocho*, sin dejar de ponderar la evolución del modelo de lengua que representan las obras lexicográficas en su sucesión.

La atención hacia la creación y la transformación del patrimonio lingüístico y cultural español forma parte de la mirada con la cual José Francisco Medina Montero y Simone Tepedino abordan, en sus respectivos artículos, el tema estudiado. En *Algunas calas en los diez primeros capítulos de la primera parte del Don Chisciotte de Franciosini: “toscanismos”, religión y germanía*, **José Francisco Medina Montero** propone un análisis traductológico contrastivo español-italiano de una parte de la obra maestra de Miguel de Cervantes, a partir de la traducción italiana de Lorenzo Franciosini de 1622. Tras contextualizar la obra original, la figura del traductor y la misma traducción al italiano, Medina Montero analiza la manera en la cual se encuentran traducidos algunos toscanismos empleados en ella, y los términos relativos a la religión y a la germanía, ya que estas palabras han dado lugar a algunos de los errores de traducción más notables, según José Francisco Medina Montero. En el análisis que propone, el autor muestra cómo dichos errores son sobre todo el fruto de la falta de precisión o bien de una interpretación errónea por parte de Franciosini.

El artículo de **Simone Tepedino**, *Discurso museístico e identidad: una comparación*, que cierra la segunda área temática de investigación, se centra en el discurso digital en el ámbito museal. Más precisamente, Simone Tepedino propone un análisis comparativo de las estrategias discursivas y lingüísticas empleadas por dos de los principales museos españoles – *El Prado* y el *Reina Sofía* – en su propia auto-representación en sendos portales web. A través de un estudio cualitativo de tipo inductivo y sincrónico, el autor muestra cómo las instituciones en cuestión poseen una propia personalidad, que promueven en particular en tres secciones del sitio web: *Historia*, *Misión* y *Educación*. El análisis de Simone Tepedino pretende por lo tanto demostrar el papel que juegan las dimensiones multimodal y pragmática en los discursos digitales analizados, los cuales, como afirma su autor, constituyen prácticas sociales, tal y como las define Fairclough (1992).

Los artículos de la tercera parte del volumen – *Transformaciones digitales en el análisis léxico, fraseológico y terminológico* – se valen de herramientas digitales para el estudio sincrónico o bien diacrónico del léxico y de la fraseología, o discuten las estrategias que las nuevas metodologías digitales proponen para enfrentarse eficazmente con los desafíos de los estudios terminológicos.

El artículo de **Andrés Ortega Garrido**, *Presencia de léxico coloquial en las entrevistas de Jordi Évole: estudio de corpus*, da paso a la tercera área temática del volumen, ya que propone un análisis del papel del léxico en el

género discursivo de la entrevista, en el contexto español. El autor se centra en el léxico de marcado carácter coloquial empleado por el periodista español Jordi Évole en el programa de televisión *Salvados*, que según Andrés Ortega Garrido promueve, gracias a este uso, la identificación del espectador. Utilizando las herramientas metodológicas de la lingüística de corpus y con el auxilio de las consideraciones teóricas propuestas desde el análisis del discurso, el estudio se basa en un conjunto homogéneo de transcripciones de veintitrés entrevistas a políticos, en su mayor parte españoles y ocasionalmente hispanoamericanos, analizadas en su vertiente léxica tanto desde el punto de vista cuantitativo, como cualitativo.

El papel central de los datos y de una base metodológica que se vale de la informática destacan en el trabajo de **Jens Fleischhauer**, *Eine datengesteuerte Untersuchung deutscher Funktionsverbgefüge. Artikelvarianz in stehen-Funktionsverbgefügen*. El autor propone un análisis, basado en corpus, de las presuntas restricciones sobre el uso de los artículos en constructos con verbo soporte en alemán. Examinando cincuenta construcciones con el verbo “stehen”, demuestra que la común afirmación sobre la limitación en el empleo de los artículos no responde al uso real, como muestran los datos. El estudio revela que dichas construcciones no tienen restricciones tan rígidas en el uso del artículo dentro del sintagma preposicional, sino que revelan preferencias variables. El autor llega a la conclusión de que las restricciones sobre el uso de los artículos no son un criterio utilizable para identificar las construcciones con verbo soporte, ni para diferenciarlas de otras construcciones predicativas alemanas. El análisis de Fleischhauer le permite evidenciar la complejidad y la variabilidad en el uso de los artículos en dichas construcciones, en oposición a las teorías precedentes, abriendo así nuevas perspectivas de investigación en este terreno.

Otros artículos, que se inscriben en la tradición lingüística francesa, contribuyen a mostrar las potencialidades de los recursos y las herramientas digitales para el análisis terminológico y fraseológico. El artículo de **Paolo Frassi**, *Entre terminologie, terminographie et humanités numériques: le projet DIACOM-fr*, presenta los retos teóricos y metodológicos de DIACOM-fr, que ha visto la creación de un banco de datos terminológico relativo al comercio internacional en perspectiva diacrónica. El trabajo de Paolo Frassi expone las etapas principales del proyecto, desde los criterios que se han tomado en consideración para la creación del corpus de referencia en relación con las diferentes finalidades y aplicaciones previstas, hasta el uso de software para la exploración del corpus, y la posibilidad de extracción automática de lexías simples y complejas. El autor expone además su reflexión sobre el modelo que hay que adoptar para representar las unidades terminológicas, basándose en las características definitorias de tipo sintáctico y semántico, en los ejes paradigmático y sintagmático. La mirada diacrónica

y la integración de aspectos didácticos constituyen otros elementos distintivos de DIACOM-fr discutidos por el autor, que ofrece además unos ejemplos de la representación gráfica de las entidades terminológicas de la base de datos.

El glosario multilingüe *Sapori locali*, el objeto de la reflexión presentada por **Chiara Preite** en su artículo titulado *Les dénominations des produits typiques de l'Émilie-Romagne. Diffusion multilingue d'une terminologie protégée*, constituye un ejemplo de las aportaciones de la transformación digital a la creación de recursos terminológicos y léxicos. El trabajo de Chiara Preite muestra cómo estos recursos pueden impulsar los estudios terminológicos, terminográficos y lexicográficos, participando al mismo tiempo en la valorización del patrimonio cultural. Dicho glosario recoge unidades terminológicas que identifican productos de la enogastronomía regional emiliana en cinco idiomas: italiano, inglés, español, francés y rumano. La autora subraya cómo los elementos léxico-culturales del glosario funcionan al mismo tiempo como testigos y motores de la materialidad local y de la cultura regional italiana, contribuyendo a la construcción de una precisa identidad social. A la presentación de las características principales del glosario y de las fichas terminológicas que lo integran sigue un análisis comparativo de los rasgos formales de algunas denominaciones protegidas en los equivalentes propuestos en francés e inglés.

El artículo de **Michela Tonti**, *Quelle souplesse pour le signe linguistique du Nom de Marque? Étude de sa variation phraséologique au prisme de la créativité du discours numérique au quotidien*, analiza desde una perspectiva semántica y fraseológica el empleo de los nombres de marca en el ámbito de un género característico del discurso digital: el blog. La autora discute los rasgos definitorios de la noción de colocación en relación con las unidades fraseológicas que incluyen un nombre de marca, resaltando la dimensión cultural y destacando el rasgo de la prominencia tanto cualitativa, como cuantitativa. De acuerdo con Michela Tonti, el uso de un nombre de marca en los discursos seleccionados adquiere un papel en cierta medida prototípico. A partir del corpus *frTenTen* (2017), la autora realiza un análisis diacrónico de algunos nombres de marca de diversos ámbitos, ahondando en el caso de *Canada Dry* y *Chocapic*. El artículo de Michela Tonti destaca la eficacia del nombre de marca utilizado en una colocación capaz de representar lo que la autora llama “activador léxico-cultural” (“*activateur lexiculturel*”).

En cambio, el artículo de **Daniela Capra**, *Usos y fraseología de la palabra “pelo” en castellano. Un estudio de corpus hasta el siglo XVII*, estudia los valores semánticos y la amplia fraseología de la palabra “pelo” en la lengua española desde sus primeras manifestaciones hasta comienzos del siglo XVII a través de la consulta de un corpus textual digital, el CORDE. De la extracción de concordancias resalta tanto la cronología de las primeras

documentaciones léxico-semánticas y fraseológicas, como la red de las relaciones que pone en conexión mutua las diferentes unidades sintagmáticas y la variación formal que tiene lugar en la diacronía. Siguiendo la tradición filológica y situándose en el terreno de la historiografía lingüística, sin olvidar la importante cuestión de las tradiciones discursivas, Daniela Capra perfila el recorrido de las acepciones semánticas de dicho vocablo y de sus numerosas locuciones y paremias; algunas de ellas todavía existen en la lengua actual, mientras que otras se han perdido tras una etapa más o menos larga de uso.

Los cinco artículos que integran la cuarta y última parte del volumen, dedicada a las *Transformaciones digitales en el análisis del discurso*, abordan el tema de las transformaciones implicadas en los discursos digitales. El artículo de **Julien Longhi** – *Explorer les mots du politique dans la transformation numérique. Analyser le lexique politique dans des contextes et selon des ressources en évolution* – muestra cómo la evolución de las metodologías y de las herramientas informatizadas permite realizar análisis socio-discursivos complejos, como respuestas a las transformaciones del discurso político en el contexto digital. La evolución de las herramientas y de las metodologías de análisis del discurso político se presentan a través de muestras de los elementos principales de diferentes proyectos y recursos realizados en el contexto de la investigación en Francia. El autor, con un estudio comparativo del léxico que caracteriza los tuits¹ de tres candidatos durante las campañas presidenciales de 2017 y 2022, muestra la evolución del estilo de los tres representantes políticos, con el auxilio de herramientas de análisis que implementan la inteligencia artificial y el *deep learning*. El estudio de Julien Longhi ilumina las relaciones entre léxico, ideologías y argumentación política.

Una reflexión metodológica sobre las aportaciones de metadatos en las exploraciones de corpus digitales con el auxilio de herramientas informáticas constituye el objeto del artículo de **Claudia Cagninelli**, *Discours numériques et approches outillées: quelques réflexions sur les apports des métadonnées*. La autora señala el papel de los metadatos en la estructuración de los corpus en formato digital como condición necesaria para llevar a cabo un análisis informatizado que tome en consideración los diferentes elementos constitutivos de los discursos digitales. Después de abordar la cuestión de las repercusiones epistemológicas, teóricas y metodológicas de las transformaciones digitales de las producciones discursivas, la autora muestra cómo los metadatos pueden materializar perspectivas y niveles de análisis

¹ Como subrayan los artículos que analizan estos discursos, a la plataforma Twitter y sus producciones, los tuits, se les ha cambiado el nombre respectivamente en X y post a partir del mes de julio de 2023. En los artículos del presente volumen, los ejemplos analizados pueden ser anteriores o posteriores a este cambio.

diferentes que, a través de un enfoque contrastivo-diferencial, promueven un estudio más completo y complejo. Finalmente, el artículo de Claudia Cagninelli propone una clasificación de los metadatos basada tanto en la naturaleza de las informaciones categorizadas, como en las características de las textualidades seleccionadas y el enfoque de análisis previsto, y la ilustra con ejemplos de aplicación a partir de tuits.

El artículo de **Marcella Palladino**, *Webbasierte Tools für die Transkription und Analyse von Reden. Hilfreiche Instrumentarien für die (Polito)Linguistik*, se centra en la solución de problemas comunes en la transcripción de discursos políticos, proponiendo el uso de herramientas web BAS (Bayerisches Archiv für Sprachsignale). La autora subraya la importancia de integrar la señal audio en el análisis lingüístico, superando la tendencia a pasarlo por alto en algunas disciplinas. Basándose en un corpus de discursos alemanes e italianos reunidos para un proyecto doctoral, el estudio presenta soluciones prácticas para investigadores en otros terrenos que, como la politolingüística, estudian discursos largos no transcritos. El enfoque metodológico que caracteriza el estudio de Palladino le permite subrayar ventajas y desventajas de estas herramientas digitales, mientras ofrece recomendaciones y buenas prácticas para un uso eficaz en la investigación lingüística.

Isabella Ferron analiza la evolución de la comunicación política alemana en las redes sociales en su artículo *Digitalisierung der politischen Kommunikation. Eine politolinguistische Analyse der Twitter/X-Kommunikation ausgewählter AfD-Politiker*, examinando un corpus de posts del partido AfD (*Alternative für Deutschland*). En este trabajo, se propone entender si la comunicación externa de algunos miembros de AfD en Twitter ha cambiado, si se orientan a formas de comunicación más tradicionales como discursos públicos o participaciones a debates de televisión, y cuáles medios lingüísticos utilizan estratégicamente para alcanzar un público más amplio. Para llegar a cumplir con este objetivo, la autora describe la comunicación política en línea, focalizándose en Twitter/X, examina la selección del corpus y el método utilizado y, finalmente, se centra en el análisis lingüístico cualitativo de una selección de posts.

Los discursos digitales de la plataforma Twitter también constituyen el objeto del análisis del artículo de **Gloria Zanella**, *Les métaphores no-vax sur Twitter. Une étude comparative français-italien*, que analiza el empleo de metáforas relativas a la vacuna contra la covid-19 en unos tuits que se adscriben a las dos posturas contrapuestas. Tras discutir el papel de las metáforas en la comunicación digital, la autora propone un estudio cualitativo de carácter comparativo-interlingüístico de algunos ejemplos de metáforas creativas antivacunas sacadas de dos corpus de tuits, uno francés y el otro italiano. El análisis de Gloria Zanella muestra diferentes formas de conflicto que caracterizan los ejemplos analizados, en particular los que oponen las dos

posturas antagonistas acerca de la vacunación y los argumentos de los movimientos antivacunas respecto a las provisiones de salud. La autora subraya además la eficacia comunicativa de las imágenes metafóricas y su reelaboración en el discurso digital.

TRANSFORMATIONEN LINGUISTISCHER ANSÄTZE IM DIGITALEN DISKURS

Die digitale Revolution zählt zu den Entwicklungen, die die sprachliche Veränderung am stärksten beeinflusst haben, da sich die Diskurspraktiken in diesem Kontext kontinuierlich weiterentwickeln. Seit dem späten 20. Jahrhundert zeigen zahlreiche Studien zu digitalen Diskursen die weitreichenden Auswirkungen neuer Technologien auf sämtliche Ebenen der sprachlichen Produktion, insbesondere im Bereich der multimodalen Kommunikation. Digitale Transformationen führen zu einer Vielzahl innovativer sprachlicher Praktiken, die mit neuartigen Online-Kommunikations- und Interaktionsinstrumenten wie Foren (Mourlhon-Dallies *et al.* 2004), Blogs und insbesondere sozialen Medien (Paveau 2017; Zappavigna 2018) sowie mit der Verwendung spezifischer kommunikativer Ressourcen und Formen wie beispielsweise Hashtags (Zappavigna 2015; Mercier 2018) oder auch Memes (Davison 2012; Simon, Wagener 2023) einhergehen. Die neuen digitalen Werkzeuge haben maßgeblich dazu beigetragen, einer wachsenden Vielfalt nationaler und internationaler Akteure Gehör zu verschaffen – institutionellen, wirtschaftlichen, sozialen und insbesondere der öffentlichen Meinung –, ermöglicht durch die Einführung des partizipativen Webs.

Die sprachlichen und diskursiven Veränderungen, die durch die Nutzung digitaler Technologien ausgelöst werden, beeinflussen auch die theoretischen und methodologischen Ansätze zur Erforschung digitaler Diskurse. Sie führen zu einer Weiterentwicklung oder Erneuerung bestehender Konzepte, insbesondere hinsichtlich der Textualität (Mayeur, Paveau 2020), des Diskurses und der diskursiven Genres (Paveau 2017). In der Tat ist die verbale Sprache nur eine von vielen Komponenten digitaler Diskurse, deren multimodaler und multimedialer Charakter das Untersuchungsfeld der linguistischen und diskursiven Studien erweitert hat. Diese richten ihren Fokus zunehmend auf die plurisemiotische und multimodale Konstruktion solcher Praktiken (Kress, van Leeuwen 2001; Paveau 2017).

Seit langem schon hebt die linguistische Forschung die Dynamik und Komplexität diskursiver Produktionen hervor und betont dabei die verschiedenen sprachlichen und außersprachlichen Elemente, die an deren kontinuierlichen und vielfältigen Transformationen beteiligt sind. Ein kontextualisierter Ansatz zur Analyse verbaler Produktionen erweist sich jedoch bei digitalen Diskursen als besonders notwendig, da die Vielzahl sprachlicher

und nichtsprachlicher Faktoren ihre spezifische Komplexität bestimmt. Die Transformationen digitaler Diskurse spiegeln zudem soziale und kulturelle Entwicklungen wider, die sich auf lexikalischer, semantischer, phraseologischer sowie generell auf textueller und diskursiver Ebene manifestieren.

Die sprachlichen und diskursiven Veränderungen im digitalen Bereich haben eine Anpassung bestehender theoretischer und methodischer Ansätze erforderlich gemacht, die zugleich von den jüngsten technologischen Entwicklungen profitieren. Der Einsatz von IT-Tools in der linguistischen und diskursiven Analyse hat zur Entwicklung neuer Forschungskonzepte beigetragen und die linguistischen Ansätze zur Sprachforschung revolutioniert. Neue Forschungsfelder und Methoden haben die computerbasierte Instrumentierung zu einem zentralen Bestandteil linguistischer Studien gemacht, wie zum Beispiel im Fall der Korpuslinguistik (Habert *et al.* 1997; Biber *et al.* 1998; Williams 2005; McEnery, Hardie 2012) und den Arbeiten im frankophonen Bereich zur Textometrie (Mayaffre *et al.* 2019) und der Textstatistik in der Datenanalyse (Lebart *et al.* 2019) sowie im hispanischen Bereich der Stilometrie (Cuéllar, Vega García-Luengos 2017-2024) und der Schaffung von abfragbaren Portalen (vgl. Carpi in diesem Band). Die Erstellung von Korpora in digitalem Format, die mit Computerwerkzeugen abrufbar sind, ist heute für viele zeitgenössische linguistische und diskursive Studien von zentraler Bedeutung, wie mehrere in diesem Band gesammelte Beiträge zeigen. Auch im Bereich der Linguistik wächst die Entwicklung digitaler Tools, die das Potenzial der künstlichen Intelligenz nutzen und sie in linguistische und diskursive Analysetools integrieren (Mayaffre, Vanni 2021; vgl. Longhi und Palladino in diesem Band).

Dieser Themenband enthält achtzehn Beiträge, die sich auf drei Sprachen konzentrieren: Französisch, Spanisch und Deutsch. Die sprachliche Vielfalt, die den Band kennzeichnet, soll durch die Gliederung der Beiträge nach vier thematischen Schwerpunkten die translinguale und transkulturelle Reflexion anregen:

1. Digitale Transformationen im Sprach- und Terminologieerwerb;
2. Digitale Transformationen der sprachlichen und diskursiven Formen des kulturellen Erbes;
3. Digitale Transformationen in der lexikalischen, phraseologischen und terminologischen Analyse;
4. Digitale Transformationen in der Diskursanalyse.

Die Beiträge in den drei genannten Sprachen sind aufeinander abgestimmt und behandeln folglich aus unterschiedlichen Perspektiven ein gemeinsames Thema. Elf der Artikel stammen aus ausgewählten Beiträgen der Konferenz *Exploring Words in the Digital Transformation: Tools and Approaches for the Study of Lexis and Phraseology in Evolving Discourse Domains*, die

2021 vom interuniversitären Forschungszentrum CLAVIER in Zusammenarbeit mit den Forschungsgruppen Lexi-Term und CAP (*Professional and Academic Communication*) organisiert wurde. Die Konferenz fand am 18. und 19. November 2021 an der Universität Modena und Reggio Emilia statt. Der vorliegende Band steht daher in engem Dialog mit den englischsprachigen Studien, die auf derselben Konferenz präsentiert und in der Ausgabe 58 von *Lingue e Linguaggi* (Bondi, Cacchiani 2023) und der Ausgabe 27 von *Status Quaestionis* (Bondi, Cacchiani 2024) veröffentlicht wurden.

Die in diesem Band gesammelten Beiträge der CLAVIER-Konferenz 2021 befassen sich mit den Auswirkungen des digitalen Wandels aus verschiedenen, aber eng miteinander verknüpften Perspektiven. Sie verdeutlichen die Wechselbeziehungen zwischen lexikalisch-phraseologischen und sozio-diskursiven Phänomenen. Zusätzlich vertiefen sieben ausgewählte Artikel diese Reflexion, indem sie die vier thematischen Hauptachsen genauer beleuchten und die untersuchten Bereiche weiter ausbauen.

Dieser Sonderband zielt darauf ab, die tiefgreifenden Veränderungen zu analysieren, die das Digitale in alltäglichen sprachlichen und diskursiven Praktiken sowie in spezialisierten Bereichen bewirkt. Zudem werden die theoretischen und methodologischen Ansätze erforscht, die zur Analyse sowohl der digitalen Kommunikation als auch der digitalisierten und/oder mittels digitaler Werkzeuge untersuchten Daten angewendet werden.

Die Beiträge beleuchten einerseits den intensiven Wandel in den sich entwickelnden digitalen Medien und die daraus resultierenden neuen Formen der Partizipation. Andererseits fokussieren sie auf die Rolle und den Beitrag digitaler Werkzeuge bei der Untersuchung von Lexikographie und Phraseologie. Darüber hinaus wird die Entwicklung in den Bereichen Lexikographie, Lexikologie, lexikalische Semantik und Diskursanalyse betrachtet. In diesen Fällen widmen sich die Beiträge sowohl der Digitalisierung historischer Werke als auch der Erstellung und Erforschung zunehmend umfangreicher und leistungsfähiger moderner Korpora.

Im ersten Forschungsschwerpunkt, der sich mit dem Thema *Digitale Transformationen im Sprach- und Terminologieerwerb* befasst, geht es um die Auswirkungen digitaler Transformationen auf die Entwicklung sprachlich-diskursiver und terminologischer Kompetenzen in vier verschiedenen Lernkontexten: in der Vorschule (Bose/Kurtenbach), an der Universität (Kaunzner/Bardeck), in der Lehrerausbildung (Gegner) und schließlich in der städtischen Bildung (Rocha Ochoa). Im diesem ersten Abschnitt wird der Beitrag digitaler Werkzeuge und Methoden zur Erforschung des Lernprozesses sowie der Einsatz digitaler Ressourcen als Lernmittel betont.

Der Artikel von **Ines Bose** und **Stephanie Kurtenbach**, „*aber das WAR AUCH n bisschen SCHÖN bei dem traum*“. *Exemplarische Studie zum*

Sprechen über inneres Erleben in Gesprächskreisen des Kindergartens, entstammt der Forschung zu Gesprächskreisen in deutschen Kindergärten. Diese zielen darauf ab, die soziale Teilhabe und Sprachentwicklung der Kinder zu fördern. Die Studie basiert auf einem Videokorpus, das die Interaktionen zwischen den Kindern dokumentiert. Die Autorinnen präsentieren eine Fallstudie über Problemlösungsstrategien bei Sechsjährigen, wobei sie sich auf deren Fähigkeit konzentrieren, innerpsychisches Erleben – Emotion, Volition und Kognition – auszudrücken. Diese Fähigkeit wird als entscheidend für die zwischenmenschliche Kommunikation erachtet. Im Einklang mit ihrem theoretischen Rahmen analysieren die Forscherinnen exemplarisch Daten aus einem Gesprächskreis. Dabei liegt der Fokus auf der Verwendung sprachlicher Mittel zum Ausdruck von Emotionen und inneren Zuständen. Besondere Beachtung findet der Erwerb dieser Fähigkeiten durch einfühlsame, von Fachkräften unterstützte Gespräche.

Ulrike A. Kaunzner und **Lisa Bardeck** präsentieren in ihrem Beitrag *Die Datenbank „Studentische Referate in der Fremdsprache Deutsch“ und ihre Anwendung in einer Pilotstudie zur rhetorischen Sprechwirkung* die Entwicklung und Nutzung einer Datenbank mit Audio- und Videoaufzeichnungen von Referaten von Studierenden des Faches Deutsch als Fremdsprache. Die Autorinnen erörtern die verschiedenen Phasen der Transkription mithilfe des OCTRA-Tools und beleuchten dessen Anwendungsmöglichkeiten sowie seine Grenzen. Eine Längsschnittstudie eines Teilkorpus, basierend auf Videoaufnahmen im Abstand von zwei Jahren, offenbart bemerkenswerte Entwicklungen in den Sprechfertigkeiten der Studierenden. Dabei werden Selbsteinschätzungen mit externen Bewertungen verglichen. Die Analyse berücksichtigt zudem Unterschiede in der Sprechwirkung zwischen Präsenz- und Online-Präsentationen. Der Fokus liegt auf dem Gesamteindruck, der Verständlichkeit, der Natürlichkeit, der wahrgenommenen Kompetenz und dem Stil, wobei verschiedene phonetische Merkmale einbezogen werden. Abschließend skizzieren die Autorinnen potenzielle Forschungsperspektiven und innovative Lehransätze.

Christian Gegner widmet sich in seinem Beitrag *Analyse der stimmlichen Leistungsfähigkeit bei Lehramtsstudierenden. Ein Vergleich zweier digital gestützter multiparametrischer Verfahren* der stimmlichen Belastung von Lehrkräften. Diese Berufsgruppe sieht sich hohen stimmlichen Anforderungen gegenüber, auf die viele Pädagogen unzureichend vorbereitet sind, was zu einer hohen Prävalenz von Stimmstörungen im Lehrberuf führt. Der Autor schlägt als Lösungsansatz ein digitales Stimmscreening vor. In seiner Studie vergleicht er zwei multiparametrische Methoden anhand einer Stichprobe von 200 Studierenden der Universität Regensburg (Deutschland). Basierend auf den Erkenntnissen dieser Untersuchung erörtert Gegner abschließend die Entwicklung einer App zur Stimmhygiene und -prophylaxe für Lehrkräfte. Dieses innovative Instrument zielt darauf ab, die stimmliche Gesundheit und Leistungsfähigkeit im pädagogischen Berufsalltag nachhaltig

zu fördern.

Der Beitrag von **Carlos Rocha Ochoa**, *Los glosarios financieros electrónicos en español y portugués como instrumentos de educación financiera en Latinoamérica*, bildet den Abschluss des ersten Forschungsbereichs dieses Bandes und leitet zugleich zum zweiten Teil über, der sich spezifisch dem Lexikon widmet. Rocha Ochoa untersucht mittels einer vergleichenden Analyse digitaler lexikografischer Instrumente deren Rolle bei der Stärkung des Bildungswesens und der Inklusion im Finanzkontext in Lateinamerika. Im Gegensatz zu vorangegangenen Beiträgen liegt der Fokus hier auf der finanziellen Alphabetisierung des Durchschnittsbürgers. Die Studie analysiert elektronische lexikografische Ressourcen nach einem funktionalen Ansatz, der die Bedürfnisse der Finanzdienstleistungsnutzer, die Zugänglichkeit und die Verarbeitungsmethodik berücksichtigt. Dabei werden auch metatextuelle Elemente wie Anweisungen und Erklärungen einbezogen. Das von Rocha Ochoa erstellte Korpus offenbart nicht nur Unterschiede zwischen den untersuchten Glossaren hinsichtlich Zugänglichkeit und Relevanz der ausgewählten Begriffe. Es stellt nach Ansicht des Autors auch eine empirische Grundlage für zukünftige Entwicklungen im Bereich spanisch- und portugiesischsprachiger lexikografischer Ressourcen als Instrumente der Finanzbildung dar.

Die Lexikographie nimmt eine Schlüsselrolle bei der Gestaltung und Transformation des kulturellen Erbes ein, dessen Sprachrohr und Zeuge sie zugleich ist. Der zweite Teil des Bandes – *Digitale Transformationen der sprachlichen und diskursiven Formen des kulturellen Erbes* – umfasst drei Beiträge, die sich diesem Themenkreis widmen. Im Zentrum steht dabei die Funktion von Wörtern als Zugang zu und Zeugnis von Identitätstransformationen sowie kulturellen Konstruktionen und Repräsentationen.

Der zweite Teil des Bandes eröffnet mit einer Abhandlung zur lexikographischen Analyse. Im Fokus steht die Vorstellung des Portals *Dicciocho* – ein innovatives Instrument für die Erforschung spanischer Wörterbücher des 18. Jahrhunderts, entwickelt von der Forschungsgruppe *Diccionarística del siglo XVIII*, der auch **Elena Carpi** angehört. In ihrem Beitrag *Dicciocho. An 18th Century Lexicographic Portal* hebt Carpi die Bedeutung des Portals hervor. Es dient als Schlüssel zur Verbreitung und Zugänglichmachung der Einträge der zweiten Ausgabe des *Diccionario de Autoridades* für die Wissenschaftsgemeinschaft – ein Werk, das lange als verschollen galt und erst kürzlich in den Archiven der Real Academia Española wiederentdeckt wurde. Carpi betont die stetige Weiterentwicklung des Portals durch die Integration von Materialien aus der ersten Ausgabe des Wörterbuchs (1726-39) sowie des *Diccionario castellano con voces de Ciencias y Artes* von Terreros y Pando. Zur Veranschaulichung der Nutzungsmöglichkeiten von *Dicciocho* wählt die Autorin Begriffe aus dem

Bereich der Kunst, insbesondere der Architektur und Malerei. Dabei beleuchtet sie die Evolution des Sprachmodells, wie es sich in der Abfolge der lexikographischen Werke manifestiert.

Die Entstehung und Veränderung des spanischen Sprach- und Kulturerbes wird aus zwei unterschiedlichen Perspektiven in den Beiträgen von Jose Francisco Medina Montero und Simone Tepedino beleuchtet. In *Algunas calas en los diez primeros capítulos de la primera parte de Don Chisciotte de Franciosini: "toscanismos", religión y germanía* präsentiert **Jose Francisco Medina Montero** eine kontrastive spanisch-italienische Translanguaging-Analyse eines Teils des *Don Chisciotte*, basierend auf Lorenzo Franciosinis italienischer Übersetzung von 1622. Nach einer Kontextualisierung untersucht er die Übersetzung ausgewählter Toskanismen sowie religiöser und *germania*-bezogener Begriffe. Diese Elemente führten laut Medina Montero zu bedeutenden Übersetzungsfehlern, die er hauptsächlich auf Ungenauigkeiten und Fehlinterpretationen zurückführt.

Den Abschluss des zweiten thematischen Bereichs bildet **Simone Tepedino** mit seinem Beitrag *Discurso museístico e identidad: una comparación*, der sich dem digitalen Diskurs im Museumskontext widmet. Tepedino vergleicht die diskursiven und sprachlichen Strategien zweier bedeutender spanischer Museen – *El Prado* und *Reina Sofia* – in ihrer Selbstdarstellung auf ihren Webportalen. Mittels einer induktiven und synchronen qualitativen Studie zeigt der Autor, wie diese Institutionen ihre eigene Persönlichkeit besitzen, und wie sie diese durch drei spezifische Bereiche auf ihren Portalen zum Ausdruck bringen: Geschichte (*Historia*), Mission (*Misión*) und Bildung (*Educación*). Tepedinos Analyse zielt darauf ab, die Rolle der multimodalen und pragmatischen Dimensionen der untersuchten digitalen Diskurse aufzuzeigen. Diese stellen, so der Autor, soziale Praktiken im Sinne der von Fairclough (1992) vorgeschlagenen Definition dar.

Der dritte Teil des Bandes – *Digitale Transformationen in der lexikalischen, phraseologischen und terminologischen Analyse* – präsentiert Beiträge, die digitale Werkzeuge für synchrone und diachrone Untersuchungen von Lexikon und Phraseologie nutzen oder Strategien erörtern, die neue digitale Methoden für effektive terminologische Studien erforderlich machen.

Andrés Ortega Garrido leitet mit *Presencia de léxico coloquial en las entrevistas de Jordi Évole: estudio de corpus* zum dritten thematischen Teil über. Der Autor konzentriert sich auf das umgangssprachlich markierte Lexikon, das der spanische Journalist Jordi Évole in der Talkshow *Salvados* verwendet, was laut Ortega Garrido die Identifikation des Publikums begünstigt. Unter Verwendung der methodischen Instrumente der Korpuslinguistik und mit Hilfe theoretischer Überlegungen, die von der Diskursanalyse vorgeschlagen werden, basiert diese Studie auf einem

homogenen Korpus, das aus der Transkription von 23 Interviews mit zumeist spanischen und einigen lateinamerikanischen Politikern besteht. Der Autor analysiert den lexikalischen Aspekt sowohl aus quantitativer als auch aus qualitativer Sicht.

Jens Fleischhauer unterstreicht in *Eine datengesteuerte Untersuchung deutscher Funktionsverbgefüge. Artikelvarianz in stehen-Funktionsverbgefügen* die zentrale Rolle von korpuslinguistischen Daten und computergestützten Methoden in der Linguistik. Mit seiner korpusbasierten Analyse von 50 Konstruktionen mit dem Verb „stehen“ stellt er die gängige Annahme in Frage, dass Funktionsverbgefüge eine eingeschränkte Artikelvarianz aufweisen. In seiner Studie zeigt Fleischhauer unterschiedliche Präferenzen in der Artikelverwendung auf und schlussfolgert, dass Restriktionen bezüglich der Artikelvarianz kein valides Kriterium für die Identifizierung von Funktionsverbgefügen oder deren Abgrenzung von anderen prädikativen Konstruktionen im Deutschen darstellen. Fleischhauers methodischer Ansatz ermöglicht es, die Komplexität und Variabilität im Gebrauch von Artikeln in Stützverbkonstruktionen hervorzuheben. Dadurch werden bisherige Annahmen angezweifelt und neue Forschungsperspektiven in diesem linguistischen Bereich eröffnet.

Weitere Artikel, die in der Tradition der französischen Linguistik verankert sind, verdeutlichen das Potenzial digitaler Ressourcen und Werkzeuge für die terminologische und phraseologische Analyse. **Paolo Frassi** beleuchtet in seinem Beitrag mit dem Titel *Entre terminologie, terminographie et humanités numériques: le projet DIACOM-fr* die theoretischen und methodischen Herausforderungen des DIACOM-fr-Projekts, das die Schaffung einer terminologischen Datenbank für den internationalen Handel aus einer diachronen Perspektive anstrebt. Er stellt die wesentlichen Phasen des Projekts vor, von den Kriterien, die bei der Erstellung des Referenzkorpus in Bezug auf die verschiedenen Zwecke und Anwendungen berücksichtigt wurden, über den Einsatz von Software zur Korpusforschung bis hin zur Möglichkeit der automatischen Extraktion sowohl einfacher als auch komplexer Begriffe. Weiter erläutert Frassi die Überlegungen zu dem Modell, das für die Darstellung der terminologischen Einheiten auf der Grundlage ihrer definierenden syntaktischen und semantischen Merkmale auf paradigmatischer und syntagmatischer Ebene herangezogen wird. Der diachrone Ansatz sowie die Integration didaktischer Aspekte sind charakteristische Merkmale von DIACOM-fr, die der Autor eingehend erörtert. Abschließend präsentiert er Beispiele für die grafische Darstellung der terminologischen Einheiten innerhalb der Datenbank.

Das mehrsprachige Glossar *Sapori locali*, das **Chiara Preite** in ihrem Beitrag *Les dénominations des produits typiques de l'Émilie-Romagne. Diffusion multilingue d'une terminologie protégée* behandelt, stellt ein Beispiel für den Einfluss der digitalen Transformation auf die Schaffung

terminologischer und lexikografischer Ressourcen dar. In ihrem Beitrag zeigt Preite, wie diese Ressourcen sowohl Impulse für die terminologischen, terminografischen und lexikografischen Studien geben als auch zur Aufwertung des kulturellen Erbes beitragen können. Das Glossar vereint Begriffe zur Bezeichnung regionaler Lebensmittel- und Weinprodukte in fünf Sprachen: Italienisch, Französisch, Englisch, Spanisch und Rumänisch. Die Autorin betont, dass die lexikokulturellen Elemente des Glossars sowohl das jeweilige Gebiet als auch die italienische Regionalkultur widerspiegeln und fördern, wodurch sie zum Aufbau einer spezifischen sozialen Identität beitragen. Nach der Einführung der wichtigsten Merkmale des Glossars und der dazugehörigen Terminologieblätter folgt eine vergleichende Analyse der formalen Merkmale einiger geschützter Namen und ihrer vorgeschlagenen französischen und englischen Pendants.

Der Artikel von **Michela Tonti** mit dem Titel *Quelle souplesse pour le signe linguistique du Nom de Marque? Étude de sa variation phraséologique au prisme de la créativité du discours numérique au quotidien* analysiert aus semantischer und phraseologischer Perspektive die Verwendung von Markennamen im Kontext eines charakteristischen Genres des digitalen Diskurses: den Blogs. Die Autorin erörtert die definierenden Merkmale des Begriffs der Kollokation in Bezug auf phraseologische Einheiten, die einen Markennamen enthalten, und hebt dabei dessen kulturelle Dimension sowie die qualitative und quantitative Salienz hervor. Tonti argumentiert, dass die Verwendung des Markennamens in den ausgewählten Diskursen eine nahezu prototypische Rolle einnimmt. Ausgehend vom frTenTen-Korpus (2017) führt die Autorin eine diachrone Analyse einiger Markennamen durch, die verschiedenen Domänen angehören, und untersucht dabei insbesondere die Fälle von *Canada Dry* und *Chocapic*. Der Artikel von Michela Tonti betont die Wirksamkeit des Markennamens, der innerhalb einer Kollokation in der Lage ist, das zu vermitteln, was die Autorin als „lexiko-kulturellen Aktivator“ (*activateur lexiculturel*) definiert.

Auch der Beitrag von **Daniela Capra** mit dem Titel *Usos y fraseología de la palabra “pelo” en castellano. Un estudio de corpus hasta el siglo XVII* untersucht die semantischen Werte und die vielfältige Phraseologie des Wortes „pelo“ in der spanischen Sprache von seinen Ursprüngen bis zum Beginn des 17. Jahrhunderts anhand eines digitalen Textkorpus. Die Extraktion von Konkordanzen liefert sowohl die Chronologie der ersten semantischen und phraseologischen Vorkommen als auch das Beziehungsgeflecht zwischen den verschiedenen syntagmatischen Einheiten sowie die formale Variation im diachronen Verlauf. Daniela Capra verfolgt einen philologischen Ansatz und bewegt sich im Bereich der Sprachgeschichtsschreibung, ohne dabei die wichtige Frage der diskursiven Traditionen zu vernachlässigen. Sie zeichnet den Weg der semantischen Bedeutungen des Wortes sowie der zahlreichen Lokutionen und Phrasen nach:

Einige davon sind in der aktuellen Sprache noch präsent, während andere nach einer mehr oder weniger langen Gebrauchszeit in den Hintergrund gedrängt wurden.

Die mit den digitalen Diskursen verbundenen Transformationen werden in den fünf Beiträgen des vierten und letzten Teils des Bandes behandelt: *Digitale Transformationen in der Diskursanalyse*. Der Artikel von **Julien Longhi**, *Explorer les mots du politique dans la transformation numérique. Analyser le lexique politique dans des contextes et selon des ressources en évolution*, zeigt, wie die Entwicklung computergestützter Methoden und Werkzeuge komplexere soziodiskursive Analysen ermöglicht, die als Antwort auf die Transformationen des politischen Diskurses im digitalen Kontext dienen. Die Entwicklung von Instrumenten und Methoden zur Analyse dieser Diskurse wird anhand der Schlüsselemente mehrerer Projekte und Instrumente veranschaulicht, die im Rahmen der französischen Forschung umgesetzt wurden. Anhand einer vergleichenden Studie des Lexikons, das die Tweets² von drei Kandidaten während der Präsidentschaftswahlen 2017 und 2022 charakterisiert, zeigt der Autor die Veränderungen im Stil dieser politischen Vertreter auf; dabei bedient er sich der Analysewerkzeuge, die die künstliche Intelligenz und *deep learning* nutzen, sowie statistischer Daten, die mit der Urheberschaft verknüpft sind. Die Studie von Julien Longhi beleuchtet somit die Beziehungen zwischen Wortschatz, Ideologien und politischer Argumentation.

Um eine methodologische Reflexion über die Rolle von Metadaten bei der Erforschung digitaler Korpora mithilfe von Computerwerkzeugen geht es im Beitrag von **Claudia Cagninelli**: *Discours numériques et approches outillées: quelques réflexions sur les apports des métadonnées*. Die Autorin betont die Bedeutung von Metadaten bei der Strukturierung digitaler Korpora als notwendige Voraussetzung für die Durchführung einer computergestützten Analyse, die in der Lage ist, die verschiedenen konstitutiven Elemente digitaler Diskurse zu berücksichtigen. Nach einer Auseinandersetzung mit den epistemologischen, theoretischen und methodologischen Auswirkungen der digitalen Transformationen in diskursiven Produktionen zeigt Cagninelli auf, wie Metadaten unterschiedliche Perspektiven und Analyseebenen eröffnen. Diese ermöglichen durch einen kontrastiv-differentiellen Ansatz eine umfassendere und komplexere Untersuchung. Abschließend schlägt die Autorin eine Klassifizierung von Metadaten vor, basierend auf der Art der kategorisierten Informationen, den Merkmalen der ausgewählten Textinformationen und dem

² Wie in einigen Artikeln, die diese digitalen Diskurse analysieren, hervorgehoben wird, wurden die Plattform Twitter und ihre Inhalte, die Tweets, ab Juli 2023 in „X“ und „Posts“ umbenannt. Die analysierten Beispiele in den Beiträgen beziehen sich sowohl auf Diskurse, die vor als auch nach diesen Änderungen entstanden sind.

beabsichtigten Analyseansatz und veranschaulicht dies mit Anwendungsbeispielen von Tweets.

Der Beitrag von **Marcella Palladino** mit dem Titel *Webbasierte Tools für die Transkription und Analyse von Reden. Hilfreiche Instrumente für die (Polito)Linguistik* befasst sich mit der Lösung gängiger Probleme bei der Transkription politischer Reden und empfiehlt den Einsatz webbasierter Tools des BAS (*Bayerisches Archiv für Sprachsignale*). Palladino betont die Notwendigkeit, Audiosignale stärker in linguistische Analysen einzubeziehen, um deren Vernachlässigung in einigen Disziplinen zu überwinden. Basierend auf einem Korpus deutscher und italienischer Reden, das im Rahmen eines Promotionsprojekts erstellt wurde, bietet die Studie praktische Lösungsansätze für Forschende, die in Bereichen wie der Politolinguistik mit langen, nicht transkribierten Reden arbeiten. Der methodische Ansatz der Studie ermöglicht es, die Vor- und Nachteile dieser digitalen Werkzeuge aufzuzeigen und Empfehlungen sowie *best practices* für deren effektiven Einsatz in der (polito)linguistischen Forschung zu formulieren.

Isabella Ferron analysiert in ihrem Beitrag *Digitalisierung der politischen Kommunikation. Eine politolinguistische Analyse der Twitter/X-Kommunikation ausgewählter AfD-Politiker* die Entwicklung der deutschen politischen Kommunikation in sozialen Medien. Dabei untersucht sie einen Korpus von Posts der Partei AfD (*Alternative für Deutschland*). Ziel des Beitrags ist es, zu ergründen, ob und wie sich die Außenkommunikation einiger AfD-Mitglieder auf Twitter/X verändert hat, ob sie auf traditionellere Kommunikationsformen wie öffentliche Reden oder Talkshow-Auftritte Bezug nehmen und welche sprachlichen Strategien sie nutzen, um ein breiteres Publikum anzusprechen. Um diese Fragestellungen zu beantworten, beschreibt Ferron die politische Online-Kommunikation mit einem Schwerpunkt auf Twitter/X, erläutert die Auswahl des Korpus und die angewandte Methode und präsentiert abschließend eine qualitative linguistische Analyse ausgewählter Posts.

Die digitalen Diskurse auf der Plattform Twitter/X werden auch im Beitrag von **Gloria Zanella** *Les métaphores no-vax sur Twitter. Une étude comparative français-italien* analysiert, in dem die Verwendung von Metaphern im Zusammenhang mit der Covid-19-Impfung in Tweets untersucht wird, die sich auf die beiden gegensätzlichen Positionen zu dieser Thematik beziehen. Nach einer Erörterung der Rolle von Metaphern in der digitalen Kommunikation schlägt die Autorin eine vergleichende, interlinguistische qualitative Studie vor, die kreative No-Vax-Metaphern aus zwei Korpora – einem französischen und einem italienischen – analysiert. Zanellas Studie zeigt verschiedene Konfliktformen auf, die die untersuchten Beispiele prägen, insbesondere den Gegensatz zwischen den beiden antagonistischen Positionen zur Frage der Impfung und den Widerstand der

No-Vax-Bewegung gegen die mit der Impfung verbundenen Gesundheitsvorschriften. Die Autorin betont zudem, wie Metaphern zur effektiven konzeptionellen Rahmung beitragen und wie sie Gegenstand einer Schichtung in diesen digitalen Diskursen sind.

TRANSFORMATIONS DES APPROCHES LINGUISTIQUES DES DISCOURS NUMÉRIQUES

La révolution numérique est l'un des événements récents qui ont le plus influencé les transformations linguistiques entraînées par l'évolution des pratiques communicationnelles dans ce contexte. Les études sur les discours numériques sont très nombreuses depuis la fin du dernier millénaire et elles ont montré les effets de l'utilisation de nouvelles technologies sur les différents niveaux des productions verbales et, plus encore, sur les productions multimodales. Les transformations produites par le numérique foisonnent dans les pratiques linguistiques liées à de nouveaux moyens de communication et d'interaction en ligne tels que les forums (Mourlhon-Dallies *et al.* 2004), les blogues et, tout particulièrement, les réseaux socio-numériques (Paveau 2017 ; Zappavigna 2018). Ces transformations découlent également du recours à des ressources et à des formes de communication spécifiques à l'instar des hashtags (Zappavigna 2015 ; Mercier 2018) ou bien des mèmes (Davison 2012 ; Simon, Wagener 2023). De nouveaux espaces de communication ont également vu le jour sur le web et ils sont accessibles à un public de plus en plus large : des acteurs nationaux et internationaux d'ordre institutionnel, économique et social, mais aussi et surtout à l'opinion publique, grâce à l'introduction du web participatif.

Les transformations linguistiques et discursives déterminées par l'utilisation des technologies numériques ont eu en conséquence des répercussions sur les perspectives théoriques et méthodologiques à adopter pour étudier les discours numériques, modifiant et/ou renouvelant les approches existantes, à partir des conceptions mêmes de la textualité (Mayeur, Paveau 2020), du discours et des genres de discours (Paveau 2017). La nature multimodale et multimédiale des discours numériques, dont le verbal n'est que l'un des éléments constitutifs, a ainsi élargi le champ d'investigation des recherches linguistiques et discursives, qui prennent en considération également la construction plurisémiotique et multimodale de telles pratiques (Kress, van Leeuwen 2001 ; Paveau 2017).

Depuis longtemps, les études en sciences du langage ont mis en évidence la vitalité et la complexité des pratiques discursives, soulignant les divers éléments linguistiques et extralinguistiques qui participent à leurs transformations constantes et variées. La prise en compte du contexte dans l'étude des productions verbales constitue donc un

aspect encore plus nécessaire pour analyser les discours numériques en raison de nombreux facteurs linguistiques et non linguistiques qui donnent lieu à une complexité spécifique. Les transformations des discours numériques présentent en outre les traces des évolutions sociales et culturelles des pratiques communicationnelles correspondantes, qui ressortent tant au niveau lexico-sémantique qu'aux niveaux phraséologique, textuel et discursif.

Bien que les transformations linguistiques et discursives produites en contexte numérique aient requis d'adapter les approches théorico-méthodologiques existantes, ces dernières ont à leur tour profité des évolutions récentes de nature technologique et numérique. Le support d'outils informatiques pour l'analyse linguistique et discursive a contribué au développement de nouvelles perspectives de recherche, révolutionnant les approches linguistiques de l'étude du langage. De nouveaux domaines et de nouvelles méthodologies considèrent d'ailleurs l'outillage informatique comme un élément fondamental de l'étude linguistique. Entre autres, on peut citer le cas de la linguistique de corpus (Habert *et al.* 1997 ; Biber *et al.* 1998 ; Williams 2005 ; McEnery, Hardie 2012), ainsi que les apports, dans le domaine francophone, de la textométrie (Mayaffre *et al.* 2019) et de la statistique textuelle dans l'analyse des données (Lebart *et al.* 2019), tout comme, dans le domaine hispanique, de la stylométrie (Cuéllar, Vega García-Luengos 2017-2024) et de la création de portails consultables (cf. Carpi dans ce numéro). La constitution de corpus en format numérique et leur exploration par des outils numériques jouent désormais un rôle de premier plan dans les recherches linguistiques contemporaines, comme en témoignent plusieurs des articles rassemblés dans ce numéro. Dans ce domaine aussi, on assiste à un développement croissant d'instruments informatiques qui se servent du potentiel de l'intelligence artificielle, en l'intégrant dans des outils d'analyse linguistique et discursive (Mayaffre, Vanni 2021 ; cf. Longhi et Palladino dans ce numéro).

Ce numéro thématique se compose de dix-huit articles qui portent sur trois langues : français, espagnol et allemand. La diversité linguistique caractérisant le numéro vise à alimenter une réflexion translinguistique et transculturelle en organisant les diverses contributions autour de quatre axes thématiques principaux :

1. transformations numériques dans l'acquisition linguistique et terminologique ;
2. transformations numériques dans les formes linguistiques et discursives du patrimoine culturel ;
3. transformations numériques dans l'analyse lexicale, phraséologique et terminologique ;
4. transformations numériques dans l'analyse du discours.

Les articles dans les trois langues sont rassemblés en fonction de la problématique partagée, développée toutefois à partir de perspectives différentes. Onze articles ont été sélectionnés à la suite du colloque intitulé *Exploring Words in the Digital Transformation. Tools and Approaches for the Study of Lexis and Phraseology in Evolving Discourse Domains*, qui a été organisé auprès de l'Université de Modène et Reggio d'Emilie les 18 et 19 novembre 2021 par le Centre Interuniversitaire de Recherche CLAVIER 2021 (Centro Interuniversitario di Ricerca CLAVIER 2021) en collaboration avec les groupes de recherche *Lexi-Term* et *CAP (Professional and Academic Communication)*. Le présent numéro se pose ainsi en dialogue étroit avec les études anglophones présentées dans le cadre de ce même colloque et constituant le numéro 58 de *Lingue e Linguaggi* (Bondi, Cacchiani 2023) et le numéro 27 de *Status Quaestionis* (Bondi, Cacchiani 2024).

Les articles ici recueillis à partir des communications présentées au colloque CLAVIER 2021 abordent les effets des transformations numériques depuis des perspectives différentes qui montrent toutefois l'interrelation entre des phénomènes lexico-phaséologiques et socio-discursifs. Sept articles sélectionnés dans un deuxième temps visent à leur tour à contribuer à la réflexion autour des quatre axes mentionnés plus haut, en élargissant les objets d'étude analysés.

L'objectif du numéro est d'analyser les transformations profondes introduites par le numérique dans les pratiques linguistiques et discursives non seulement ordinaires mais aussi dans des domaines spécialisés. Il en va de même pour les perspectives théoriques et méthodologiques adoptées pour analyser tant la communication qui a lieu dans l'environnement numérique, que les données numériques et/ou explorées par le biais d'outils informatiques. D'une part, les contributions abordent les changements profonds affectant les médias numériques en évolution, avec de nouvelles formes de participation et de contribution. De l'autre, l'attention est dirigée vers le rôle et l'apport des outils informatiques dans l'étude du lexique et de la phraséologie, mais aussi au développement de la lexicographie, de la lexicologie, de la sémantique lexicale et de l'analyse du discours. Dans ces cas, les contributions portent aussi bien sur la numérisation de textes anciens que sur la création et exploration de corpus récents de plus en plus larges et performants.

Dans le premier axe de recherche consacré aux *Transformations numériques dans l'acquisition linguistique et terminologique*, les études s'intéressent aux impacts des transformations numériques sur l'acquisition de compétences linguistico-discursives et terminologiques dans quatre contextes d'apprentissage différents : le milieu préscolaire (Bose/Kurtenbach), le milieu universitaire (Kaunzner/Bardeck), la formation des enseignants (Gegner) ainsi que le milieu de l'éducation citoyenne (Rocho Ochoa). Les études

regroupées dans cet axe mettent en évidence l'apport des outils informatiques et des méthodologies pour l'étude du processus d'apprentissage, de même que l'utilisation de ressources numériques comme moyen d'apprentissage.

L'article de **Ines Bose** et **Stephanie Kurtenbach**, „*aber das WAR AUCH n bisschen SCHÖN bei dem traum*“. *Exemplarische Studie zum Sprechen über inneres Erleben in Gesprächskreisen des Kindergartens*, est issu d'une recherche sur les cercles de conversation dans un jardin d'enfance allemand, dont l'objectif est de promouvoir la participation sociale et le développement linguistique. L'étude prend appui sur un corpus de vidéos qui documente les interactions entre les enfants. Les auteures présentent une étude de cas concernant les stratégies de résolution de problèmes chez les enfants de six ans, en se concentrant en particulier sur leur capacité à exprimer l'expérience psychologique intérieure (émotions, volition, cognition), qui s'avère cruciale dans la communication interpersonnelle. Conformément aux présupposés théoriques, les données d'un cercle de conversation sont analysées à titre d'exemple, mettant en évidence l'emploi de moyens linguistiques pour parler d'émotions et d'états intérieurs vécus lors d'une conversation empathique gérée par des professionnels.

Dans leur contribution, intitulée *Die Datenbank „Studentische Referate in der Fremdsprache Deutsch“ und ihre Anwendung in einer Pilotstudie zur rhetorischen Sprechwirkung*, **Ulrike A. Kaunzner** et **Lisa Bardeck** présentent et traitent l'élaboration d'une base de données constituée d'enregistrements audio et vidéo de présentations réalisées par des étudiants d'allemand comme langue étrangère. Elles réfléchissent sur les différentes étapes de la transcription effectuée via l'outil OCTRA, en mettant en lumière ses possibilités d'exploitation ainsi que ses limites. Une étude longitudinale d'un sous-corpus d'enregistrements vidéo à distance capturés au long de deux ans révèle des changements intéressants dans les capacités oratoires des étudiants et elle dresse une comparaison entre les autoévaluations et les évaluations externes. L'analyse porte ensuite sur les différences entre les présentations réalisées en présentiel et à distance, en se focalisant sur l'impression générale, la clarté, le naturel, la compétence perçue et le style, sur la base de différentes caractéristiques phonétiques. Des perspectives de recherche et des approches didactiques font également l'objet de la réflexion.

Dans *Analyse der stimmlichen Leistungsfähigkeit bei Lehramtsstudierenden. Ein Vergleich zweier digital gestützter multiparametrischer Verfahren*, **Christian Gegner** attire l'attention sur la voix des enseignants, en tant que catégorie professionnelle soumise à des fortes exigences vocales. Néanmoins, nombre d'enseignants n'ont pas de préparation adéquate pour faire face à ces exigences, ce qui entraîne une prévalence remarquable de troubles vocales. L'article propose un filtrage vocal numérique et compare deux méthodes multiparamètres à partir d'un échantillon de 200 étudiants de l'Université de Regensburg (Allemagne). La

conclusion aborde le développement d'une application pour l'hygiène et la prévention vocale pour les professionnels de l'enseignement.

La contribution de **Carlos Rocha Ochoa**, *Los glosarios financieros electrónicos en español y portugués como instrumentos de educación financiera en Latinoamérica*, conclut le premier axe de recherche du volume, favorisant la transition vers la deuxième partie qui se concentre de manière plus spécifique sur le lexique. À travers une analyse comparative d'outils lexicographiques numériques, Carlos Rocha Ochoa s'intéresse au rôle qu'ils jouent dans la promotion de l'éducation et de l'inclusion dans le secteur financier dans l'Amérique latine. À la différence des articles précédents, l'auteur dirige ici l'attention sur l'alphabétisation du citoyen ordinaire au sein d'un domaine spécialisé. Les ressources lexicographiques numériques sur lesquelles porte l'étude de Carlos Rocha Ochoa sont analysées depuis une approche fonctionnelle, en prenant en considération les exigences des utilisateurs des services financiers, l'accessibilité et la méthodologie d'élaboration, y compris les éléments métatextuels tels que les instructions et les explications d'utilisation. L'article démontre la disparité qui caractérise les glossaires analysés en ce qui concerne l'accessibilité et la pertinence des termes sélectionnés. Le corpus constitué par Carlos Rocha Ochoa représente, d'après l'auteur, une base empirique pour des développements futurs dans le domaine des ressources lexicographiques en espagnol et en portugais en tant que moyens d'éducation financière.

La lexicographie joue toutefois un rôle tout aussi décisif dans la création et la transformation du patrimoine culturel dont la langue est le porte-voix et le témoin. Le rôle des mots comme une voie d'accès et une forme d'attestation des transformations des constructions et des représentations identitaires et culturelles constitue le sujet thématique central de la deuxième partie du volume – *Transformations numériques dans les formes linguistiques et discursives du patrimoine culturel* – qui recueille trois contributions.

L'analyse lexicographique ouvre la deuxième partie du numéro avec la présentation d'une ressource dédiée aux dictionnaires espagnols du XVIII^e siècle, le portail lexicographique *Dicciocho*, créé par le groupe de recherche *Grupo de investigación Diccioarística del siglo XVIII*, dont fait partie **Elena Carpi**. Dans son article, *Dicciocho. An 18th Century Lexicographic Portal*, Elena Carpi met en relief l'intérêt de ce portail en tant qu'outil de diffusion et d'accès pour la communauté scientifique à la nomenclature de la deuxième édition de *Autoridades*, un dictionnaire qui a été longtemps considéré comme perdu et qui a été retrouvé récemment dans les archives de la *Real Academia Española*. L'auteure souligne en outre l'expansion progressive du portail grâce à l'intégration de matériaux issus de la première édition du dictionnaire (1726-39) et du *Diccionario castellano con voces de Ciencias y Artes* de Terreros y Pando. Carpi illustre enfin l'utilisation du portail *Dicciocho* à

travers des mots relatifs au domaine artistique, et notamment à l'architecture et à la peinture, en traitant de l'évolution du modèle linguistique représenté par les ouvrages lexicographiques dans leur succession.

La création et la transformation du patrimoine linguistique et culturel espagnol constituent également l'objet d'étude des articles de José Francisco Medina Montero et de Simone Tepedino, bien que les deux adoptent deux perspectives diverses. Dans *Algunas calas en los diez primeros capítulos de la primera parte del Don Chisciotte de Franciosini: "toscanismos", religión y germanía*, **José Francisco Medina Montero** propose une analyse traductologique contrastive espagnol-italien d'une partie du *Don Chisciotte* d'après la traduction italienne de Lorenzo Franciosini de 1622. L'auteur analyse la traduction de quelques toscanismes et de lemmes relatifs à la religion et à la *germanía* puisqu'il estime que ces éléments ont donné lieu à certaines des erreurs de traduction les plus importantes. Dans son analyse, José Francisco Medina Montero montre comment ces erreurs résultent notamment d'imprécisions et d'interprétations erronées.

La deuxième section thématique se termine par la contribution de **Simone Tepedino**, *Discurso museístico e identidad: una comparación*, qui porte sur le discours numérique en contexte muséal. Plus précisément, Simone Tepedino réalise une analyse comparative des stratégies discursives et linguistiques employées par deux des principaux musées espagnols – *El Prado* et *Reina Sofía* – dans le cadre de leur autoreprésentation sur les sites web respectifs. Par le biais d'une étude qualitative inductive et synchronique, l'auteur montre comment ces institutions se caractérisent par une personnalité propre qu'elles promeuvent à travers trois sections spécifiques du site : *Historia*, *Misión* et *Educación*. L'analyse de Simone Tepedino vise donc à démontrer le rôle des dimensions multimodale et pragmatique des discours numériques étudiés, qui constituent – comme l'affirme l'auteur – des pratiques sociales conformément à la définition proposée par Fairclough (1992).

Les articles de la troisième partie du volume – *Transformations numériques dans l'analyse lexicale, phraséologique et terminologique* – se servent d'outils numériques pour l'étude synchronique et diachronique du lexique et de la phraséologie, ou bien ils traitent des défis posés par les nouvelles méthodologies numériques et des façons dont ils sont relevés de manière efficace dans les études terminologiques.

La contribution de **Andrés Ortega Garrido**, *Presencia de léxico coloquial en las entrevistas de Jordi Évole: estudio de corpus*, favorise la transition vers le troisième volet thématique du numéro, proposant une analyse du rôle du lexique dans le genre de l'interview dans le contexte espagnol. L'auteur s'intéresse notamment au vocabulaire familier utilisé par le journaliste espagnol Jordi Évole lors du talk-show *Salvados*, dans la mesure où le recours à ces formes lexicales marquées facilite, d'après

l’auteur, l’identification du public. En utilisant les outils méthodologiques de la linguistique de corpus et en s’appuyant sur des considérations théoriques proposées par l’analyse du discours, l’étude de Andrés Ortega Garrido repose sur un ensemble homogène de transcriptions de vingt-trois interviews de représentants politiques, majoritairement espagnols et occasionnellement latino-américains, afin d’analyser leur aspect lexical, selon une approche quantitative et qualitative.

La centralité des données et d’une approche méthodologique outillée ressort également de l’article de **Jens Fleischhauer**, *Eine datengesteuerte Untersuchung deutscher Funktionsverbgefüge. Artikelvarianz in stehen-Funktionsverbgefügen*. L’auteur réalise une étude sur corpus des restrictions présumées sur l’emploi des articles au sein des constructions à verbe support en allemand. L’analyse d’une cinquantaine de constructions avec le verbe *stehen* montre que cette affirmation courante dans la littérature allemande n’est pas étayée par les données linguistiques attestées. L’étude illustre que ces constructions n’ont pas de restrictions relatives à l’emploi de l’article à l’intérieur du syntagme prépositionnel, bien qu’elles présentent des préférences variables. L’auteur conclut que les restrictions sur les articles ne constituent pas un critère valide pour identifier les constructions à verbe support ou pour les distinguer d’autres constructions prédicatives en allemand. L’approche méthodologique adoptée permet à l’auteur de mettre en évidence la complexité et la variabilité de l’utilisation des articles dans ces constructions, ce qui remet en question les hypothèses antérieures à ce sujet et ouvre de nouvelles perspectives de recherche.

D’autres articles, s’inscrivant dans la tradition linguistique française, mettent en relief à leur tour les potentialités de ressources et d’outils numériques pour l’analyse terminologique et phraséologique. La contribution de **Paolo Frassi**, *Entre terminologie, terminographie et humanités numériques : le projet DIACOM-fr*, aborde les défis théoriques et méthodologiques de DIACOM-fr, qui a donné lieu à une base de données terminologique concernant le commerce international dans une perspective diachronique. L’article de Paolo Frassi présente les étapes principales du projet, des critères pris en considération pour la création du corpus de référence en fonction des différentes finalités et applications prévues, à l’utilisation de logiciels pour l’exploration de ce dernier, à la possibilité d’extraire automatiquement des termes simples et complexes. L’auteur illustre ensuite la réflexion sur le modèle à adopter pour la représentation des unités terminologiques, compte tenu de leurs caractéristiques définitoires de nature syntaxique et sémantique, aux niveaux paradigmatique et syntagmatique. L’approche diachronique et l’intégration d’aspects didactiques représentent des éléments distinctifs de DIACOM-fr qui sont discutés par Paolo Frassi, avant de proposer des exemples de la représentation graphique d’entités terminologiques de la base de données.

Le glossaire multilingue *Sapori locali*, qui fait l'objet de la réflexion présentée par **Chiara Preite** dans sa contribution intitulée *Les dénominations des produits typiques de l'Émilie-Romagne. Diffusion multilingue d'une terminologie protégée*, est un exemple des apports de la transformation numérique à la création de ressources terminologiques et lexicographiques. L'article de Chiara Preite montre comment ces ressources peuvent stimuler les études terminologiques, terminographiques et lexicographiques, en valorisant en même temps le patrimoine culturel. Ce glossaire se compose de termes qui désignent des produits œnogastronomiques régionaux dans cinq langues : italien, français, anglais, espagnol et roumain. L'auteure souligne comment les éléments lexicoculturels du glossaire témoignent et promeuvent aussi bien le territoire que la culture régionale italienne, contribuant à la construction d'une identité sociale spécifique. La présentation des caractéristiques principales du glossaire et de ses fiches terminologiques est suivie d'une analyse comparative des caractéristiques formelles de quelques dénominations protégées dans les termes correspondants en français et en anglais.

L'article de **Michela Tonti**, *Quelle souplesse pour le signe linguistique du Nom de Marque ? Étude de sa variation phraséologique au prisme de la créativité du discours numérique au quotidien*, analyse dans une perspective sémantique et phraséologique l'utilisation des noms de marque au sein d'un genre caractéristique du discours numérique : le blogue. L'auteure discute les traits définitoires de la notion de *collocation* en relation avec les unités phraséologiques incluant un nom de marque. Elle en met en évidence la dimension culturelle, en soulignant le trait de la saillance non seulement qualitative mais aussi quantitative. D'après Michela Tonti, le recours au nom de marque dans les discours analysés joue en effet un rôle presque prototypique. À partir du corpus *frTenTen* (2017), l'auteure réalise une étude diachronique de quelques noms de marque issus de domaines divers, en approfondissant notamment le cas de *Canada Dry* et de *Chocapic*. L'article de Michela Tonti met en relief l'efficacité du nom de marque employé au sein d'une collocation représentant ce que l'auteure définit comme un « activateur lexicoculturel ».

De même, la contribution de **Daniela Capra**, *Usos y fraseología de la palabra "pelo" en castellano. Un estudio de corpus hasta el siglo XVII*, analyse les valeurs sémantiques et la phraséologie du mot « pelo » dans la langue espagnole depuis ses origines jusqu'au début du XVII^e siècle, à travers la consultation d'un corpus textuel numérique. L'extraction des concordances révèle la chronologie des premières occurrences sémantiques et phraséologiques, le réseau des relations reliant les diverses unités syntagmatiques ainsi que les variations formelles se produisant dans la diachronie. En adoptant une approche philologique qui se situe dans le champ de l'historiographie linguistique, sans oublier la question importante des

traditions discursives, Daniela Capra retrace le parcours accompli par les acceptions sémantiques du mot et de nombreuses locutions et parémies. Certaines de ces dernières sont encore utilisées dans la langue courante, alors que d'autres sont sorties de l'usage après une période d'emploi plus ou moins longue.

Les transformations caractérisant l'analyse des discours numériques sont abordées par les cinq contributions de la quatrième et dernière partie thématique du volume qui porte sur les *Transformations numériques dans l'analyse du discours*. L'article de **Julien Longhi** – *Explorer les mots du politique dans la transformation numérique. Analyser le lexique politique dans des contextes et selon des ressources en évolution* – montre comment l'évolution des méthodologies et des outils informatiques permet de réaliser des analyses socio-discursives plus complexes en réponse aux transformations du discours politique en contexte numérique. Les évolutions des outils et des méthodologiques pour l'analyse du discours politique sont illustrées d'abord en présentant les aspects fondamentaux de plusieurs projets et outils réalisés dans le cadre de la recherche française. À travers une étude comparative du vocabulaire caractérisant les tweets³ de trois candidats aux élections présidentielles de 2017 et de 2022, l'auteur montre ensuite les transformations des styles des trois représentants politiques en se servant d'outils d'analyse qui implémentent l'intelligence artificielle et le *deep learning*. La contribution de Julien Longhi met ainsi en relief les relations entre la dimension lexicale, les idéologies et l'argumentation politique.

Une réflexion méthodologie sur les apports des métadonnées dans l'exploration de corpus numériques par le biais d'outils informatiques fait l'objet de l'article de **Claudia Cagninelli**, *Discours numériques et approches outillées : quelques réflexions sur les apports des métadonnées*. L'auteure souligne le rôle des métadonnées dans la structuration de corpus en format numérique en tant que condition nécessaire pour réaliser une analyse outillée qui prend en considération les divers éléments constitutifs des discours numériques. Après avoir abordé les répercussions épistémologiques, théoriques et méthodologiques des transformations numériques dans les productions discursives, l'auteure montre comment les métadonnées peuvent matérialiser différentes perspectives et divers niveaux d'analyse des données qui, dans le cadre d'une approche contrastive-différentielle, favorisent une étude plus complète et complexe. L'article de Claudia Cagninelli propose enfin une typologie des métadonnées en fonction de la nature des informations catégorisées, ainsi que des caractéristiques des textualités

³ Comme le soulignent les divers articles analysant ces discours numériques, la plateforme Twitter et ses productions, les tweets, ont été respectivement rebaptisés X et posts à partir de juillet 2023. Selon les différentes contributions, les exemples analysés correspondent à des discours produits avant ou après ces changements.

sélectionnées et de l'approche d'analyse envisagée, en l'illustrant par des exemples d'application à partir de tweets.

La contribution de **Marcella Palladino**, ayant pour titre *Webbasierte Tools für die Transkription und Analyse von Reden. Hilfreiche Instrumentarien für die (Polito)Linguistik*, concerne la résolution de problèmes courants dans la transcription de discours politiques, en proposant de recourir à des outils web BAS (*Bayerisches Archiv für Sprachsignale*). L'auteure souligne l'importance de prendre en compte le signal audio dans l'analyse linguistique, dépassant la tendance de certaines disciplines à le négliger. Se basant sur un corpus de discours allemands et italiens constitué dans le cadre d'un projet de doctorat, l'étude de Marcella Palladino présente des solutions pratiques pour les chercheurs appartenant à d'autres domaines de recherche, tels que la linguistique appliquée au domaine politique, qui traitent de longs discours non transcrits. L'approche méthodologique caractérisant l'étude de Palladino permet à l'auteure d'éclairer les avantages et les inconvénients de ces outils numériques. Elle propose en outre des recommandations et de bonnes pratiques pour leur utilisation efficace dans la recherche linguistique relative au discours politique.

Dans sa contribution *Digitalisierung der politischen Kommunikation. Eine politolinguistische Analyse der Twitter/X-Kommunikation ausgewählter AfD-Politiker*, **Isabella Ferron** s'intéresse à l'évolution de la communication politique allemande sur les réseaux socionumériques, en analysant un corpus de posts du parti AfD (*Alternative für Deutschland*). L'auteure se propose de comprendre si la communication externe de quelques membres de l'AfD sur Twitter/X a changé, en reprenant des formes de communication plus traditionnelles à l'instar des discours publics ou des participations à des talk-show. Elle étudie également les stratégies et les moyens linguistiques utilisés afin d'atteindre un public de followers plus large. Pour ce faire, l'auteure décrit la communication politique en ligne en se concentrant sur Twitter/X. Après avoir traité de la sélection du corpus et de la méthode adoptée, l'auteure réalise une analyse linguistique qualitative d'une sélection de posts.

Les discours numériques de la plateforme Twitter/X font également l'objet de l'analyse de l'article de **Gloria Zanella**, *Les métaphores no-vax sur Twitter. Une étude comparative français-italien*. Dans ce cas, l'attention se dirige sur le recours à des métaphores relatives à la vaccination contre le Covid-19 au sein de tweets relevant des deux positions qui s'opposent autour de ce sujet. Après avoir discuté du rôle des métaphores dans la communication numérique, l'auteure propose une étude qualitative d'ordre contrastif-interlinguistique de quelques exemples de métaphores créatives « no-vax » issues de deux corpus de tweets, l'un français et l'autre italien. L'étude de Gloria Zanella montre différentes formes de conflit qui caractérisent les exemples analysés, opposant en particulier les deux positions antagonistes sur l'administration ou pas du vaccin, ainsi que l'opposition de

la position qualifiée de « no-vax » aux dispositions sanitaires liées à la vaccination. L’auteure met en outre en évidence comment les métaphores contribuent à véhiculer un cadrage (*framing*) conceptuel efficace et comment elles sont stratifiées au sein de ces discours numériques.

Direcciones de correo de las editoras / Adressen der Herausgeberinnen / Adresses des directrices: claudia.cagninelli@unimi.it; daniela.capra@unige.it;
ulrike.kaunzner@unimore.it; chiara.preite@unimore.it

Bibliografía / Literatur / Références bibliographiques

- Biber D., Conrad S. and Reppen R. 1998, *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge University Press.
- Bondi M. and Cacchiani S. (eds) 2023, *Lingue e Linguaggi 58: Positioning and Identity in Digital Discourse*.
- Bondi M. and Cacchiani S. (eds) Forthcoming 2024, *Status Quaestionis 27: Language Variation: Perspectives on Lexis and Phraseology*.
- Cuéllar Á. y Vega García-Luengos G. (2017-2024), *Proyecto ETSO*. Recurso web <http://etso.es/>
- Davison P. 2012, *The Language of Internet Memes*, in Mandiberg M. (ed.), *The Social Media Reader*, New York University Press, New York, USA, pp. 120-134.
- Fairclough N. 1992, *Discourse and Social Change*, Polity Press, Cambridge.
- Habert B., Nazarenko A. et Salem A. 1997, *Les linguistiques de corpus*, Armand Colin, Paris.
- Kress G.R. and Leeuwen T. van 2001, *Multimodal Discourse: The Modes and Media of Contemporary Communication*, Arnold, London; Oxford University Press, New York.
- Lebart L., Pincemin B. et Poudat C. 2019, *Analyse des données textuelles*, Presses de l'Université du Québec, Québec.
- Mayaffre D., Pincemin B. et Poudat C. 2019, *Explorer, mesurer, contextualiser. Quelques apports de la textométrie à l'analyse de discours*, in « Langue française » 203 [3], pp. 101-115.
- Mayaffre D. et Vanni L. (éds) 2021, *L'intelligence artificielle des textes : des algorithmes à l'interprétation*, Honoré Champion éditeur, Paris.
- Mayeur I. et Paveau M.-A. (éds) 2020, *Corela [En ligne] HS-33 : Textuel, textiel. Repenser la textualité numérique*, <https://doi.org/10.4000/corela.11734>
- McEnery T. and Hardie A. 2012, *Corpus Linguistics: Method, Theory and Practice*, Cambridge University Press, Cambridge.
- Mercier A. 2018, *Hashtags : tactiques de partages et de commentaires d'informations*, in Mercier A. et Pignard-Cheynel N. (éds), *#Info : commenter et partager l'actualité sur Twitter et Facebook*, Éditions de la Maison des sciences de l'homme, Paris, pp. 87-129.
- Mourlhon-Dallies F., Rakotonaelina F. et Reboul-Touré S. (éds) 2004, *Les Carnets du Cediscor 8 : Les discours de l'internet. Nouveaux corpus, nouveaux modèles ?*. <https://doi.org/10.4000/cediscor.68>
- Paveau M.-A. 2017, *L'analyse du discours numérique. Dictionnaire des formes et des pratiques*, Hermann, Paris.
- Simon J. et Wagener A. (éds) 2023, *Semen 54 : Approches discursives des mèmes en politique*, Presses Universitaires de Franche-Comté, Besançon.
- Williams G. (éd.) 2005, *La linguistique de corpus*, Presses Universitaires de Rennes, Rennes.
- Zappavigna M. 2015, *Searchable Talk: The Linguistic Functions of Hashtags*, in "Social Semiotics" 25 [3], pp. 274-291.
- Zappavigna M. 2018, *Searchable Talk: Hashtags and Social Media Metadiscourse*, Bloomsbury Academic, an imprint of Bloomsbury Plc, London/New York.

„aber das WAR AUCH n bisschen SCHÖN bei dem traum“ EXEMPLARISCHE STUDIE ZUM SPRECHEN ÜBER INNERES ERLEBEN IN GESPRÄCHSKREISEN DES KINDERGARTENS

INES BOSE, STEPHANIE KURTENBACH
MARTIN-LUTHER-UNIVERSITÄT HALLE-WITTENBERG

Abstract – This paper stems from a larger research project that analyses a specific format of conversation circles in a German kindergarten with respect to its capacity to promote aspects of societal participation as well as language development. We present a case study of problem-solving strategies among preschoolers (6 years) in conversation circles that took place regularly in this kindergarten. We focus on children’s ability to speak about inner psychological experience (emotion, volition, cognition). This ability is a very essential in the field of interpersonal communication. It presupposes that a child can perceive and reflect on his own inner experience and that of other persons. Having laid out our theoretical starting points as well as the current research, we shall introduce our data, an excerpt from a conversation circle about a girl’s desire, for which the group is looking for solutions. Methodologically we employ a conversation analytic framework. Subsequently, the occurring linguistic terms for inner psychological experience (inner state terms) are listed, categorized and interpreted. It turns out that professionals and children have a variety of verbal, paraverbal, and nonverbal resources at their disposal to talk about inner state. Many formulations refer to the areas of emotion/ affect and evaluation/ graduation. The participants refer to each other in their utterances and adopt terms of inner experience from each other. Through such an empathic conversation dynamic and supported by the professionals, children can acquire the ability to talk about inner experience in a context-appropriate way.

Keywords: conversation circles; speaking about inner psychological state; verbal, paraverbal, and nonverbal resources; interaction.

1. Einleitung

Im Kindergarten erzählt die sechsjährige Pauline von ihrem schlimmen Traum über einen Streit mit ihrer Freundin Julia. Dieser Traum sei „aber auch ein bisschen schön“ gewesen, denn sie hätten sich wieder vertragen. Mit dieser ambivalenten Bewertung versprachlicht das Mädchen ihr innerpsychisches Erleben. Die Fähigkeit, über eigene und fremde Gefühle, Wünsche, Bedürfnisse, Absichten und Gedanken zu sprechen, aber auch diejenigen

anderer zu deuten und darüber ins Gespräch zu kommen, ist ein wichtiger Schritt in der kindlichen Kommunikationsentwicklung. In diesem Zusammenhang ist das Gesprächskreisprojekt „Stolpersteine und Wunschsterne“ eines deutschen Kindergartens interessant. Mit Unterstützung von pädagogischen Fachkräften diskutieren hier fünf- bis sechsjährige Kinder über ihre Wünsche und Probleme und suchen nach Ideen für deren Verwirklichungen oder Lösungen. Die Gesprächskreise bieten den Kindern dabei auch die Gelegenheit, über ihre Gefühle und Bedürfnisse zu sprechen. Der vorliegende Beitrag untersucht, welche verbalen, para- und nonverbalen Ressourcen die Kinder hierfür nutzen, inwiefern sie im Gesprächsprozess auf die Äußerungen anderer Kinder zu innerem Erleben eingehen und wie sie von den pädagogischen Fachkräften dabei unterstützt werden.

2. Sprechen über inneres Erleben im Vorschulalter

Die Fähigkeit, über Innerpsychisches sprechen zu können, setzt voraus, dass ein Kind eigenes inneres Erleben und dasjenige anderer Personen wahrnehmen und darüber reflektieren kann. Damit steht diese Fähigkeit in engem Zusammenhang mit der Entwicklung der Theory of Mind (Bretherton, Beeghly 1982; Röska-Hardy 2011; Klann-Delius 2015). Denn anders als für konkret wahrnehmbare Gegenstände und Vorgänge aus der dinglichen Umwelt gibt es für inneres Erleben in der Regel keine unmittelbaren außersprachlichen Referenten. Insofern weisen Wörter und Ausdrücke mit innerpsychischem Inhalt einen höheren Grad an Abstraktheit auf und stellen einen wichtigen Schritt auf dem Weg zum Erwerb von abstrakten Begriffen dar (Bretherton *et al.* 1986; Kauschke 2012a; Bahn *et al.* 2017). Durch die Versprachlichung werden innere Zustände bewusst zugänglich und können dadurch besser kontrolliert und geteilt werden (Nelson 1996; Kauschke 2012a; Streubel *et al.* 2020 und Grosse *et al.* 2021).

Das Vokabular innerer Zustände und Prozesse besteht aus mehreren Bereichen. Die Versprachlichung von Emotionen und Affekten gehört ebenso dazu wie die von Wahrnehmungen (z. B. Geruchs- und Tastempfindungen), physiologischen Zuständen (z. B. Hunger, Durst), Volitionen (z. B. Wünsche, Bedürfnisse, Fähigkeiten), Kognitionen (z. B. Wissen, Denken, Unsicherheit) und moralischen Bewertungen sowie Selbstverpflichtungen (Bretherton, Beeghly 1982, S. 907; Kristen *et al.* 2012, S. 634). Allerdings erscheint eine klare, eindeutige Kategorisierung sehr schwierig. Deshalb plädiert zum Beispiel Klann-Delius (2015, S. 138 ff.) für eine integrative, interpersonelle und soziale Perspektive auf das komplexe Verhältnis von Affekt, Emotion, Kognition und Sprache in der kindlichen Entwicklung und weist zudem auf die Verschränkung von verbalen Formulierungen mit para- und nonverbalen Ausdrucksformen hin.

Das Sprechen über Inneres beginnt ab dem Alter von etwa 18 Monaten und wird in einer sprachenübergreifenden Reihenfolge vom Wahrnehmbaren zum Vorgestellten erworben. Das ergaben Längsschnitt- und Querschnittstudien sowie Elternbefragungen (z. B. Bretherton, Beeghly 1982; Bretherton *et al.* 1986; Klann-Delius, Kauschke 1996; Wilden, Russell 2008; Kristen *et al.* 2014; Grosse *et al.* 2021). Demnach verfügen Kinder ab dem dritten Lebensjahr über Wörter für Körperempfindungen („Hunger“), Wünsche und Absichten („wollen“), für Basisemotionen („traurig“) und deren Ausdruck („weinen“) sowie für kognitive Zustände und Prozesse („denken“). Der Anteil von Formulierungen über inneres Erleben am Gesamtwortschatz von zwei- bis dreijährigen Kindern ist im Vergleich zu Wörtern mit Bezug auf Konkretes, äußerlich Wahrnehmbares gering; das gilt insbesondere für Kognitionswörter (Aldridge, Wood 1997; Kauschke 2012a). Vom dritten bis zum fünften Lebensjahr nimmt das Emotionsvokabular sprunghaft zu (Klann-Delius 2012; Petermann, Wiedebuch 2016; Grosse *et al.* 2021) und verdoppelt sich bis zum Alter von 11 Jahren alle zwei Jahre (Baron-Cohen *et al.* 2010). Vorschulkinder können eigene und fremde Gefühlszustände aus den Bereichen Schmerz, Müdigkeit, Abscheu, Liebe, Moral benennen, über vergangene und zukünftige Emotionen sowie deren Konsequenzen sprechen und imaginierte Emotionen in fiktiven Kontexten spielerisch ausdrücken (Bretherton *et al.* 1986; Klann-Delius 2012). Das Vokabular über Inneres differenziert sich aus und wird komplexer, so können zum Beispiel Ambivalenzen bzw. Inkongruenzen zwischen innerem Erleben und seinem Ausdruck erfasst und besprochen werden (Aldridge, Wood 1997). Kauschke (2019, S. 264) schlussfolgert aus den empirischen Studien, „dass für Kinder im Vorschulalter bereits eine lexikalische Klasse ‚Gefühle‘ existiert und dass sich anfänglich breiter gefasste Emotionskategorien allmählich verengen, bis sich die Bedeutung von Emotionsbegriffen der zielsprachlichen Bedeutung annähert.“

Der Erwerb des innerpsychischen Vokabulars, seine Erweiterung und Differenzierung werden durch die Frequenz und die Spezifik des elterlichen Inputs beeinflusst (z. B. Grosse *et al.* 2021 für das Emotionsvokabular). Gerade in der Gesprächskommunikation mit vertrauten Bezugspersonen über Gefühlszustände können die Kinder die Bedeutung innerpsychischer Begriffe besonders gut erschließen, die Angemessenheit von Gefühlen hinterfragen und konkretes Verhalten evaluieren (Kauschke 2012a, S. 9).

Die meisten empirischen Studien richten sich auf den Erwerb des Vokabulars für inneres Erleben über verschiedene Altersgruppen hinweg. Sie basieren meist entweder auf Auskünften von erwachsenen Bezugspersonen (Elternfragebögen mit vorgegebenen Wortlisten) und Kindern (Aufzählung möglichst vieler Formulierungen nach Aufforderung) oder auf kontrolliert-experimentellen, elizitierenden Situationen, in denen Kinder anhand von

Stimuli zu Formulierungen über inneres Erleben ermuntert werden (zur kritischen Würdigung dieser Methoden z. B. Aldridge, Wood 1997; Streubel *et al.* 2020).

Demgegenüber konzentrieren wir uns in der folgenden Analyse auf eine bestimmte Altersgruppe (sechsjährige Kinder) und ihren interaktiven Umgang mit innerem Erleben in authentischen Gesprächssituationen, also solchen, die nicht zum Zwecke wissenschaftlicher Untersuchungen hergestellt wurden.

3. Datengrundlage und Analyseverfahren

3.1. Gesprächskreisprojekt „Stolpersteine und Wunschsterne“

Das Gesprächskreisprojekt „Stolpersteine und Wunschsterne“ in einem Kindergarten in Halle (Saale) zielt auf gesellschaftliche Teilhabe und Mitbestimmung. Die Kinder sind eingeladen, den Kindergartenalltag aktiv mitzugestalten. Dazu gehört, dass sie in diesem Gesprächskreisprojekt ausschließlich eigene Themen diskutieren. Gemeinsam mit zwei frühpädagogischen Fachkräften sprechen die Kinder der beiden ältesten Gruppen über Probleme („Stolpersteine“), für die sie eine Lösung suchen, oder über Wünsche („Wunschsterne“), deren Erfüllung sie anstreben. Das Projekt ist nicht nur hinsichtlich der Entwicklung von Partizipation und Gesprächsfähigkeit interessant, sondern auch im Hinblick auf das Sprechen über Inneres. Es verlangt von den frühpädagogischen Fachkräften eine einfühlsame Gesprächsleitung und stellt hohe kognitive und kommunikative Anforderungen an die Kinder. Sie üben Reflexions- und Analysefähigkeiten wie Erzählen, Erklären, Schildern, Berichten und Begründen, aber auch Fähigkeiten wie Empathie und Perspektivenübernahme (ausführlich Bose, Kurtenbach 2019).

3.2. Datenkorpus und Datenaufbereitung

Die Gesprächskreise wurden über drei Perioden hinweg videografiert (2014/15, 2015/16, 2018/19; Gesamtumfang: ca. 11 h). Verwendet wurden zwei Kameras (Standkamera auf die Gruppe gerichtet, Handkamera auf die jeweils sprechende Person). Die Datenschutzbestimmungen wurden beachtet, Fachkräfte und Eltern haben der Videografie und der Auswertung der Daten in anonymisierter Form schriftlich zugestimmt. Die Videoaufnahmen wurden als Feintranskript nach dem System GAT 2 (Selting *et al.* 2009) transkribiert, die Legende der verwendeten Transkriptionszeichen befindet sich am Ende des Beitrags. Alle Namen wurden nach einem Maskierungsschlüssel anonymisiert.

Als Grundlage für eine effiziente analytische Arbeit wurde eine Datenbank zur Inventarisierung, Verknüpfung der Gesprächs- und Metadaten,

Kategorisierung und gezielten Suche (z. B. zum Fallvergleich) angelegt. Diese Datenbank ermöglicht einen ökonomischen Überblick und Umgang mit dem sehr umfangreichen Datenmaterial (vgl. Deppermann 2008, S. 31) und erleichtert das Auffinden typischer Fälle (Typenbildung, Ankerbeispiele, Selektion bzw. Kollektion einschlägiger oder typischer Gesprächssequenzen) und analyserelevanter Stellen. Die Datenbank ist in einer permanenten Weiterentwicklung und Ausdifferenzierung begriffen. Aktuell verbindet sie Informationen zur Gesprächsinventarisierung (Inventarnummer, Aufnahmedatum, -ort, -zeit, Beteiligte bzw. Siglen, Dauer, Themen etc.) mit den Ergebnissen qualitativer und erster quantitativer Analysen. Nach und nach werden die anonymisierten Aufnahmen zusammengeführt mit den kompletten Basis-Transkripten der Gespräche sowie den Feintranskripten ausgewählter Gesprächsphasen und den zeitmarkenbezogenen Annotationen. Diese Zusammenschau ermöglicht makroskopische Vergleiche, zum Beispiel von Themen, Themenentwicklungen und Sprechanteilen, aber auch die Suche nach Vergleichspassagen sowie mikroskopische Analysen. Das Auflösungsniveau der Datenbank ist bewusst niedrig angesetzt, damit ein Überblick über das Gesamtkorpus und ein offener erster Zugang gewahrt bleiben.

Für die Thematik des vorliegenden Beitrages wurde das Datenkorpus der Periode 2015/16 ausgewählt, da es sich um das größte Korpus der drei Perioden handelt (Umfang 4 h 20 min, Gesamtwortzahl ca. 37.500 Wörter). Über zehn Monate hinweg wurden 16 Gesprächskreise videografiert.

3.3. Forschungsfragen und Analyseverfahren

Das Datenmaterial wirft eine ganze Reihe von Forschungsfragen auf, die im Zusammenhang mit der Versprachlichung von inneren Wahrnehmungen stehen. Anhand des Korpus wurden bereits Aspekte der Entwicklung kindlicher Gesprächsfähigkeit bearbeitet, zum Beispiel die Förderung von Partizipation und Gesprächsfähigkeit (Bose, Kurtenbach 2019), die Argumentationsförderung durch pädagogische Fachkräfte (Kurtenbach *et al.* 2019; Bose *et al.* 2020) und die argumentative Funktion von Second Stories (Wala *et al.* under review). Der Fokus der folgenden Untersuchung liegt auf den Ressourcen, die die Kinder und Fachkräfte nutzen, wenn sie im Gesprächsprozess über eigenes und fremdes inneres Erleben sprechen. Das Forschungsinteresse richtet sich darauf,

- welche Formulierungen im Korpus auftauchen und wie sie im Kontext der Anliegenbehandlung interaktiv verwendet werden,
- welche Ausdrucksmittel darüber hinaus die Gesprächsbeteiligten verwenden, um miteinander über Inneres zu sprechen.

Im Hinblick auf die Thematisierung innerer Zustände und Prozesse wird im Folgenden zunächst ein Überblick über die Anliegen der Kinder in der Periode

2015/16 gegeben. In einer ersten Analyse werden anhand der Bearbeitung eines kindlichen Anliegens die genutzten Ressourcen für das Sprechen über Inneres exemplarisch rekonstruiert. Hier geht es um die kontextuelle Einbettung der Bezeichnungen für inneres Erleben und den interaktiven Umgang damit. In einer zweiten Analyse werden dann, wie in den meisten Erwerbsstudien üblich, die sprachlichen Bezeichnungen für innerpsychische Zustände und Prozesse aufgelistet, kategorisiert und interpretiert sowie mit dem Vorkommen im Gesamtkorpus 2015/16 kontrastiert.

4. Ergebnisse

4.1. Anliegen der Kinder

In den 16 Gesprächskreisen der Periode 2015/16 bringen die Kinder insgesamt 109 Anliegen ein, 42 ordnen sie und die Fachkräfte der Kategorie „Stolpersteine“ zu und 67 der Kategorie „Wunschsterne“. Pro Gesprächskreis werden – abhängig von der Gesprächsdauer und der Teilnehmerzahl – zwischen drei und 12 Themen besprochen. Die Anliegen der Kinder verändern sich während des Projektes (ausführlich Bose, Kurtenbach 2019, S. 122 ff.). Viele Anliegen werden nur ganz kurz behandelt, zum Beispiel Ausflugs- oder Fantasiewünsche (mit allen Kindern in den Wald gehen, einen Wohnwagen für den Kindergarten kaufen). Länger wird ein Thema dann bearbeitet, wenn eine besondere Ernsthaftigkeit bzw. Betroffenheit deutlich wird (ausführlich ebd.). Bis zum neunten Gesprächskreis sprechen die Kinder überwiegend Wünsche und Probleme an, die sich auf den Kindergarten beziehen. Ab dem zehnten Gesprächskreis bringen sie gelegentlich auch schwerwiegende persönliche Probleme ein. Offensichtlich fassen die Kinder im Verlauf der Periode großes Vertrauen: Sie erleben, dass ihre Sorgen ernst genommen werden, und können sich darauf verlassen, dass die anderen Kinder und die Fachkräfte damit sensibel umgehen. Manche Stolpersteine beschäftigen die Kinder über einen längeren Zeitraum und werden in mehreren Gesprächskreisen angesprochen – zum Beispiel Streit mit anderen Kindern, Drängeln auf der Treppe, aber auch Ärger in der Familie. Die Kinder sind während der Behandlung solcher Themen emotional stark involviert. Manchmal kommt es vor, dass mit einem Wunschstern ein sehr persönliches Anliegen angesprochen wird, wie das folgende Analysebeispiel zeigt.

4.2. Sprechen über Inneres anhand der Bearbeitung eines Wunschsterns

Das Beispiel stammt aus dem elften Gesprächskreis, der insgesamt 16 Minuten dauert. Beteiligt sind sieben sechsjährige Kinder und zwei pädagogische

Fachkräfte. Die Kinder sitzen im Halbkreis den pädagogischen Fachkräften (pFK1 und pFK2; nicht im Bild) gegenüber, siehe Abbildung 1.

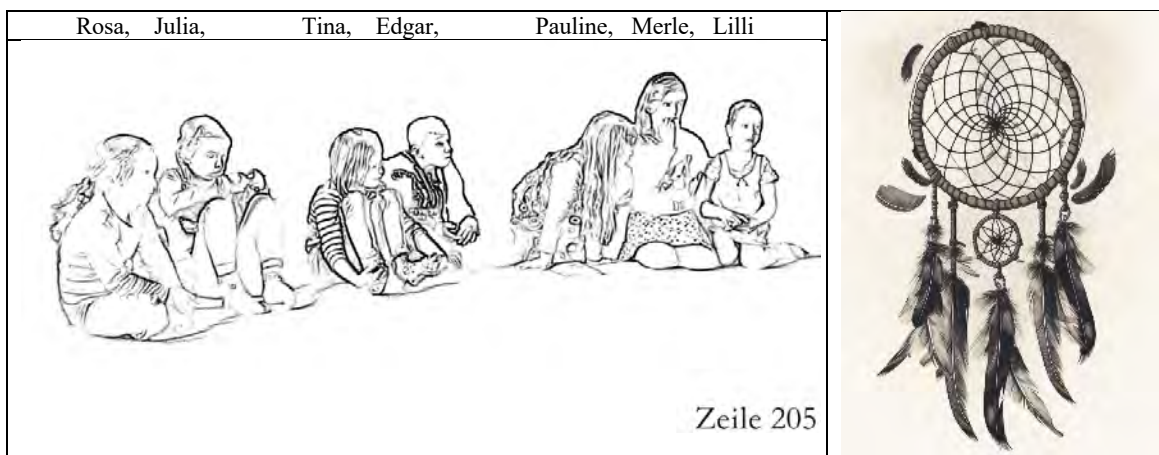


Abbildung 1
Standbild aus dem 11. Gesprächskreis (Transkriptzeile 205).

Abbildung 2
Traumfänger.

Es geht um den Wunschstern von Lilli: *dass ich bald n RAUMwegFÄNger beKOMme.*||. Lilli hat Artikulationsschwierigkeiten – gemeint ist ein so genannter „Traumfänger“ (s. Abbildung 2). Er wird am Bett aufgehängt und soll den Schlaf verbessern, indem er nur die guten Träume durch sein Netz gehen lässt, wohingegen die schlechten Träume im Netz hängen bleiben und im Morgenlicht verschwinden. Das Thema wird in Minute 09:38 gesetzt und zügig bearbeitet (2 min 44 sec): Die Gründe für den Wunsch werden erörtert und es werden Möglichkeiten zu seiner Erfüllung entworfen.

Die Analyse der Bearbeitung von Lillis Wunschstern folgt einem sequenziellen Vorgehen, ohne dass die Sequenzialität als rein chronologische Abfolge verstanden wird. Stattdessen wird die Entwicklung des Themas anhand der Bearbeitungsschritte der Gruppe betrachtet (ausführlich zu diesem Vorgehen Bose, Kurtenbach 2019; Bose *et al.* 2020). Im Transkript sind die Formulierungen, mit denen über Inneres gesprochen wird, jeweils unterstrichen.

Sequenz 1: Lilli wünscht sich einen Traumfänger (Themenformulierung)

192	pFK2:	SO.
193		LILLi;
194		[noch EINen WUNSCHstern.]
195	pFK1:	[(gib mal HER...)]
196	Pauline:	ich hab AUCH noch nen WUNSCHste:rn.
197	Rosa:	ich [hab AUCH noch n WUNSCHstern]
198	pFK2:	[<u>kannst</u> du(.) <u>KANNST</u> du GLEICH noch] SAgen;
199		jetzt ERSTmal EINmal die LILLi.

200		<<@Lilli>(.)was HAST du für einen <u>WUNSCHstern.</u> >
201	Lilli:	(.)<<p, len, mon, :-> dass ich bald n <u>RAUMwegFÄNger beKOMme;</u> >
202	pFK2:	<<mf> ein <u>↑RAUM`fänger.</u> >
203	Lilli:	<<p, len, mon, :-> ein <u>RAUMfänger</u> für meine <u>schlimme RÄUme</u> - >
204	pFK2:	<<p> für deine <u>schlimmen TRÄUme.</u> >
205	Lilli:	((nickt))
206	pFK2:	<<mf> <u>oKAY.</u> >

Der Erzieher (pFK2) gibt Lilli, die einen Wunschstern gezogen hatte, das Wort (Z. 192-194 und 198-200). Lilli formuliert ihren Wunsch: *dass ich bald n RAUMwegFÄNger beKOMme;*|| (Z. 201). Sie wirkt verlegen, denn sie spricht sehr leise, langsam, monoton, mit einem angedeuteten starren Lächeln und schaut leicht zur Seite. Der Erzieher wiederholt prosodisch expressiver (melodischer, lauter) und morphologisch leicht verändert (korrigierend): *ein ↑RAUM`fänger.*|| (Z. 202). Lilli übernimmt die korrigierte Bezeichnung des Erziehers und bestimmt die Funktion des Traumfängers, weiterhin monoton, leise sprechend und verlegen lächelnd: *ein RAUMfänger für meine schlimme RÄUme;*|| (Z. 203). Der Erzieher wiederholt diese Funktionsbestimmung von Lilli ebenfalls leise, aber phonetisch korrekt: *für deine schlimmen TRÄUme.*|| (Z. 204) und Lilli nickt (Z. 205). Mit einem abschließenden mitfühlend gesprochenen *oKAY.*|| (Z. 206) signalisiert er, dass das Thema nun gesetzt ist.

Ihre innere Betroffenheit drückt Lilli vor allem nonverbal und paraverbal aus. Neben der emotiven Metapher „Wunschstern“, die im Gesprächskreisprojekt zentral und damit etabliert ist, verwendet Lilli die Metapher „Traum(weg)fänger“ und die emotive Bezeichnung „schlimme Träume“ (Z. 201 und 203), die vom Erzieher wiederholt werden (Z. 202 und 204). Diese Ausdrucksformen prägen die Atmosphäre in der Gruppe von nun an: Die anderen Kinder reagieren ernst, fast betroffen und zeigen ihre Anteilnahme, indem sie sich wiederholt vorbeugen und Lilli anblicken sowie eigene Erfahrungen oder Lösungsvorschläge einbringen (siehe dazu die nächsten Sequenzen). Zunächst aber wird das Thema spezifiziert.

Sequenz 2: Lilli erklärt, was der Traumfänger macht (Themenspezifizierung)

207	Pauline:	[ich HAB schon einen;]
208	pFK2	[und was MACHT der dann] mit deinen <u>schlimmen träumen-</u> >
209	Lilli:	<<mf, mon> der FÄNGT dann meine <u>schlimmen räume AUF;</u> >
210	Pauline:	oder ich <u>KÖNNte</u> ja(.)mal [MEInen]
211	pFK2:	[(<<@Paul.,p> wart]

		WARte ganz kurz, >
212		<<@Lilli> und DANN hast du <u>nur noch SCHÖne träume.</u> >
213	Lilli:	((nickt) mhm,)
214	Pauline:	[oder ich]
215	pFK2:	[U:ND (.)] KAUFST du dir den denn SELber?
216	Lilli:	NEE: MAMa kauft mir den;
217	Pauline:	sie <u>KANN</u> [ja auch-]
218	Lilli:	[oder PAPA.]

Wie im Gesprächskreisformat üblich, bittet der Erzieher Lilli, das Thema näher auszuführen, indem er fragt: *und was MACHT der dann mit deinen schlimmen träumen-*|| (Z. 208). Lilli erklärt die Funktionsweise des Traumfängers: *der FÄNGT dann meine schlimmen räume AUF;*|| (Z. 209). Sie spricht immer noch monoton, aber etwas energischer (lauter) und ohne das angedeutete starre Lächeln. Der Erzieher folgert: *und DANN hast du nur noch SCHÖne träume.*|| (Z. 212). Damit bringt er eine antonymische Formulierung zu den schlimmen Träumen ein. Lilli nickt: *mhm,*| (Z. 21). Darüber hinaus fragt der Erzieher, ob sich Lilli den Traumfänger selbst kauft (Z. 215). Lilli verneint und sagt, dass ihre Eltern ihn kaufen (Z. 215-218).

Mit Bestätigungen und Nachfragen versprachlicht der Erzieher seine Anteilnahme und sorgt dafür, dass Lillis Anliegen spezifiziert wird.

Sequenz 3: Pauline bietet an, ihren Traumfänger an Lilli auszuborgen (Lösungsvorschlag und -bewertung)

Überlappend zur Themenspezifizierung hatte Pauline bereits in der zweiten Sequenz mehrfach versucht, Beiträge einzubringen: Sie hatte ihren eigenen Traumfänger erwähnt (Z. 207) und begonnen, Ideen zur Erfüllung von Lillis Wunsch zu formulieren (Z. 210, 214). Der Erzieher hatte das bisher entweder ignoriert und stattdessen das Gespräch mit Lilli weitergeführt (Z. 208, 215, 217) oder Pauline gebeten abzuwarten (Z. 211). Nun aber kommt sie zu Wort.

219	Pauline:	(.)ich <u>KANN</u> ihr ja auch mal MEInen,
220		(.)soLANge sie noch keinen HAT ausborgen weil ich HAB schon einen;
221		nen ROsanen.
222	PFK2:	oh DAS ist ja <u>NETT</u> von dir.
223	Pauline:	ich <u>KÖNNte</u> ihn ja mal mit in n KINdergarten nehmen und dann <u>KÖNNte</u> ich ihr sie-
224		°h ihr solange sie keinen HAT,
225		°h AUSborgen und wenn SIE dann einen°h HAT,
226		<u>könnte</u> sie mir den ja dann WIEdergeben.
227	PFK2:	<<@Lilli> WÄre das was? >
228	Lilli:	((nickend) mhm,)

229	PFK2:	DAS <u>finde</u> ich <u>aber</u> <u>SEHR</u> <u>nett</u> von dir pauline.
230	Pauline:	WEIL ICH(.)hab !EH! nie <u>schlimme</u> TRÄUme.
231	PFK2:	DAS <u>finde</u> ich <u>ja</u> [KLASse.]
232	PFK1:	[du hast] <u>nur</u> (.) <u>SCHÖne</u> <u>träume</u> ?
233	Pauline:	((nickt))
234	Merle:	<<pp,mon> (ich brauch) AUCH mal <u>traumfänger</u> . >
235	PFK2:	<<@Merle> du hast AUCH häufig <u>ALPträume</u> ; >
236	Merle:	<<mf> JA. >
237	PFK2:	<<zustimmend> mhm>
238	PFK2:	<<@Pauline> <u>ei</u> dann MACHen wir das [mal so;]
239	Julia:	[<u>ja</u> oKAY ich AUCH.]
240	PFK2:	<u>kannst</u> du den mal MITnehmen,
241		wenn du den sowieso nicht brauchst-
242		und dann LEIHST du den: LILLi aus.
243	Pauline:	BIS sie keinen <u>TRAUMfänger</u> -
244		(.)soLANge sie keinen <u>TRAUMfänger</u> [HAT.]
245	PFK2:	[das <u>find</u>] ich <u>ja</u> <u>wirklich</u> <u>NETT</u> von dir.

Pauline hat einen Lösungsvorschlag für Lillis Problem: Sie bietet an, ihren eigenen Traumfänger an Lilli auszuborgen (Z. 219-221, 223-226). Während Pauline ihr Angebot entwickelt, blickt Lilli sie mehrfach an und lächelt. Auch andere Kinder beugen sich vor und schauen Pauline an. Sie hat konkrete Vorstellungen davon, wie ihr Vorschlag umzusetzen wäre: Sie könnte den Traumfänger in den Kindergarten mitbringen und sobald Lilli dann selbst einen Traumfänger hat, *könnte sie mir den ja dann WIEdergeben.*|| (Z. 226). Sie verwendet wiederholt den Konjunktiv II und stellt damit Lilli die Annahme des Vorschlags frei. Außerdem nennt sie zwei Gründe für ihr Angebot: Sie hat selbst einen Traumfänger (*weil ich HAB schon einen;| nen ROsanen.*|| Z. 220-221) und sie hat *!EH! nie schlimme TRÄUme.*|| (Z. 230). Sie greift also wie vorher der Erzieher Lillis Formulierung von den „schlimmen Träumen“ auf. Der Erzieher nimmt den Irrealis von Pauline auf und fragt Lilli, ob sie das Angebot annehmen will: *WÄre das was?*|| (Z. 227). Lilli nickt: *mhm,* (Z. 228).

Der Erzieher würdigt das Angebot von Pauline mehrfach und sehr nachdrücklich: *oh DAS ist ja NETT von dir.*|| (Z. 222); *DAS finde ich aber SEHR nett von dir pauline.*|| (Z. 229); *DAS finde ich ja KLASse.*|| (Z. 231); *das find ich ja wirklich NETT von dir.* || (Z. 245). Er verwendet unterschiedliche Formulierungen zur Wertschätzung: die Interjektion („oh“), explizite Bewertungen („(finde ich) nett“, „klasse“), modifizierende Partikeln („ja“, „ja wirklich“, „aber sehr“), jeweils mit freundlichem Sprechausdruck (legatohafte

Melodiebewegungen, heller Stimmklang, melodisch-emphatische Akzente). Modellhaft bietet er also mehrere wertschätzende Ausdrucksmöglichkeiten an.

Sequenz 4: Julia erzählt von ihren Alpträumen

Bereits in der dritten Sequenz hatte Merle sehr leise gesagt, sie brauche auch mal einen Traumfänger (Z. 234):

234	Merle:	<<pp,mon> (ich brauch) AUCH mal <u>traumfänger.</u> >
235	PFK2:	<<@Merle; p> du hast AUCH häufig <u>ALPträume;</u> >
236	Merle:	<<mf> JA. >
237	PFK2:	<<zustimmend> mhm >
238	PFK2:	<<@Pauline, mf> <u>ei</u> dann MACHen wir das [mal so; >]
239	Julia:	[ja <u>oKAY</u> ich AUCH.]

Der Erzieher hatte nachgefragt: *du hast AUCH häufig ALPträume;* || (Z. 235) und damit ein verstärkendes Synonym für „schlimme Träume“ eingeführt. Auch Julia hatte kurz angedeutet, dass sie Alpträume habe: *ja oKAY ich AUCH.* || (Z. 239). Jetzt, nachdem der Wunsch von Lilli bearbeitet und mit Paulines Angebot eine vorübergehende Lösung gefunden ist, berichtet Julia über ihre Alpträume.

246	Julia:	<<p, mon> PFK2 <u>WEIßT</u> du ich hab ganz AUCH ganz oft-
247		ähm °h <u>ALTträume</u> mein bruder AUCH;
248		deswegen ist er auch die GANze nacht immer AU:F. >
249	PFK2:	<<@Julia> <u>oKAY,</u> >
250	Julia:	<<mf, mon> und DER der WECKT mich dann immer und SAGT dass °h dass dass er <u>altTRÄUme</u> hat. >
251	PFK2:	[<<zustimmend> mh>]
252	PFK1:	[<<@Julia>habt IHR] denn einen <u>TRAUMfänger-</u>
253	Julia:	((kopfschüttelnd) mm.)
254		<<mf, mon> und MAMA und PAPA <u>wollen</u> mir auch keinen KAUFen. >
255	PFK2:	was MACHST denn du wenn der TORsten dich AUFweckt?
256		was MACHST du dann,
257	Julia:	<<mf, mon> dann: SAG ich <u>KANNST</u> du wieder in dein ZIMmer gehen-
258	Julia:	°h und °h und dann GEHT er wieder in sein ZIMmer. >
259	PFK2:	<<zustimmend> mh> <u>oKAY.</u>

Julia redet den Erzieher mit einem Kognitionsverb an: *PFK2 WEIßT du* (Z. 246). Dann wiederholt sie noch einmal, dass sie auch ganz oft Alpträume habe (Z. 246) und ihr Bruder ebenfalls. Sie wirkt betroffen, denn sie spricht über die gesamte Story hinweg ähnlich wie Lilli bei ihrem Wunschstern (Z. 201): sehr leise, langsam, monoton, mit einem starren, verlegen wirkenden Lächeln. Die anderen Kinder beugen sich vor und blicken sie an, während sie die Folgen ausführt: *deswegen ist er auch die GANze nacht immer AU:F.*|| (Z. 248) *und DER der WECKT mich dann immer und SAGT dass °h dass dass er altTRÄUme hat.*|| (Z. 250). Der Erzieher bestätigt Julias Ausführungen mitfühlend (Z. 249: *oKAY*,|| und Z. 251: *mh.*) und die Erzieherin fragt, ob sie und ihr Bruder einen Traumfänger haben (Z. 252). Das verneint Julia und ergänzt: *und Mama und Papa wollen mir auch keinen KAUFen.*|| (Z. 254). Der Erzieher fragt Julia, was sie mache, wenn ihr Bruder sie aufwecke (Z. 255-256), und sie antwortet: *dann: SAG ich KANNST du wieder in dein ZIMmer gehen-|°h und °h und dann GEHT er wieder in sein ZIMmer.*|| (Z. 257-258), der Erzieher bestätigt das wiederum (Z. 259). Weiter bearbeitet wird der Wunsch von Julia nach einem Traumfänger in der Gruppe nicht.

Julia greift das von Lilli eingebrachte Thema der „schlimmen Träume“ bzw. „Alpträume“ auf – sowohl inhaltlich, indem sie von ähnlichen Erfahrungen berichtet, als auch durch einen ähnlich betroffen klingenden Sprechausdruck.

Sequenz 5: Lilli und Pauline erzählen (fiktive) schlimme Träume

Nun erzählt Lilli wie zur Illustration ihrer schlimmen Träume einen vermutlich fiktiven Traum. Der Traum betrifft Mara, die zur Gruppe gehört, aber nicht anwesend ist.

260	Lilli:	<<:-),mf> (aber) ich hab (heute) was <u>!SCHLIM!mes geträumt; ></u>
261	PFK2:	<<p> was HAST du denn <u>geträumt?</u> >
262	Lilli:	<< f, :-)> dass MArA sich in son-
263		(.)da dass MArA sich in so n> <<„jammernd“> <u>!BIEST!</u> verwandelt `HA:T <u>!O:´O::H!, ></u>
264	Pauline:	und ICH hab °h [<u>geträu</u>]
265	PFK2:	[<<ff> !UN!sre !´MA!`ra? >]
266	Lilli:	<<„jammernd“, :-)> JA:- >
267	Pauline:	[und ICH hab]
268	PFK2:	[<<„erschrocken“, :-),f> °hhh das <u>KANN</u>] ich mir <u>ja gar nicht !^VOR!stellen; ></u>
269	Merle:	((schlägt „erschrocken“ Hand vor aufgerissenen Mund))
270	Pauline:	und ICH hab
271		[(.)HEUTE <u>geträumt</u>]

272	Lilli:	[<<:-),ruf> und ma MArA sah ganz !GRU:!selig aus; >
273	Pauline:	[und ICH hatte HEUTE <u>geträumt</u> ,]
274	Pauline:	°h DAS aber-
275		°h das WA:R(.)AUCH °h n <u>bisschen</u> <u>SCHÖN</u> bei dem traum,
276		dass ICH und <<:-)> JULia uns <u>geSTRITten</u> haben aber DANN haben wir uns wieder <u>verTRA:gen</u> . >
277	PFK2:	oKAY.
278	Lilli:	<<„jammernd“ ,:-)> OH das war so ^GRU:selig. >

Lilli beginnt mit der Bemerkung, sie habe heute etwas Schlimmes geträumt (Z. 260), aber sie lächelt dabei offen. Mehrere andere Kinder beugen sich vor und blicken sie an. Auf die leise Nachfrage des Erziehers, was sie denn geträumt habe (Z. 261), führt Lilli weiterhin lächelnd aus: *dass MArA sich in son- (.) da dass MArA sich in so n !'BIEST! verwandelt 'HA:T !O:'O:H!*,|| (Z. 262-263). Sie malt den fiktiven Schrecken über die Verwandlung von Mara prosodisch aufwändig aus, insbesondere die Bezeichnung „so ein Biest“ und die kommentierende Interjektion „oh“ (laut und hoch mit großen steigend-fallenden melodischen Bewegungen, emphatischen Akzenten, starken Lautdehnungen). Der Erzieher spielt die Fiktion mit, indem er ebenfalls sehr laut und übertrieben erstaunt nachfragt: *!UN!sre !'MA!'ra?*|| (Z. 265), auf Lillis Bestätigung (Z. 266) gespielt erschrocken einatmet und wiederum prosodisch expressiv, geradezu exklamatorisch ausruft: *°hhh das KANN ich mir ja gar nicht !^VOR!stellen;*|| (Z. 268). Merle beteiligt sich an der Ausmalung des fiktiv-gruseligen Geschehens nonverbal, indem sie die Hand vor den aufgerissenen Mund schlägt (Z. 269). Lilli ruft anschließend weiterhin lächelnd aus: *und ma MArA sah ganz !GRU:!selig aus;*|| (Z. 272; sehr hoch, laut, Lautdehnung, emphatischer Akzent) und wiederholt später noch einmal ebenso expressiv: *OH das war so ^GRU:selig.*|| (Z. 278). Möglicherweise bezieht sich Lilli inhaltlich und performativ auf Rezeptionserfahrungen mit Märchen und Geschichten wie zum Beispiel „Von einem, der auszog, das Gruseln zu lernen“ oder „Die Schöne und das Biest“.

Während Lillis Erzählung hatte Pauline zu Lilli geschaut und bereits dreimal zu einem eigenen Beitrag angesetzt (Z. 264, 267 und 270-271). Nun berichtet auch sie von einem Traum am heutigen Tag. Es geht um einen Streit zwischen ihr und der anwesenden Julia. Zu Beginn charakterisiert Pauline ihren Traum: *aber das WA:R (.) AUCH °h n bisschen SCHÖN bei dem traum,*| (Z. 275) und führt dann lächelnd aus: *dass ICH und JULia uns geSTRITten haben aber DANN haben wir uns wieder verTRA:gen.*|| (Z. 276). Auch Julia lächelt während dieser Erzählung und blickt Pauline kurz an. Der Erzieher reagiert freundlich bestätigend auf Paulines Erzählung: *oKAY.*|| (Z. 277).

In der bisherigen Bearbeitung von Lillis Wunschstern (Sequenzen 1 bis 4) wurden die Inhalte der schlimmen Träume und ihre Auswirkungen auf das innere Erleben der Kinder verbal nicht behandelt. Interessant ist, dass das nun Lilli und Pauline ohne Ermutigung durch die Fachkräfte tun, indem sie über vermutlich fiktive Träume berichten. Die beiden Kinder benennen jeweils den Inhalt des Traums und schildern ihr inneres Erleben. Sie verwenden dafür emotive Formulierungen, Metaphern und Graduierungen: Lilli sagt, sie habe „was Schlimmes geträumt“, Mara habe sich „(in ein) Biest (verwandelt)“ und alles sei „ganz / so gruselig“. Pauline kennzeichnet ihren Streit-Versöhnungs-Traum als ambivalent, als „auch ein bisschen schön“. Zur expressiven Ausmalung des fiktiv-gruseligen Geschehens in Lillis Traum nutzen die Kinder und der Erzieher auch para- und nonverbale Mittel. Insbesondere bei Lilli ist der Kontrast zum prosodisch reduzierten, geradezu verlegenen Sprechen bei der Themensetzung (erste Sequenz) sehr deutlich.

Sequenz 6: Merle hat eine Lösung für ihr Problem und der Erzieher schließt das Thema ab (Lösungsbestätigung)

279	Merle:	(-)ich frag MEIne eltern <u>AUCH mal</u> ob ich n
280		[<u>TRAUMfänger</u> haben kann.]
281	PFK2:	[<<@Lilli> na da BIN ich <u>ja</u> ma jetzt] <u>geSPANNT.</u>
282		wenn die pauLine ihren <u>traumfänger</u> (.)MITnimmt,
283		ob DEIne träume dann <u>SCHÖner</u> werden.
284		(.)da BIN ich <u>ja</u> mal <u>gespannt</u> was DU <u>erZÄHLST.</u>
285		(.)oKAY. >
286		DANN hatte jetzt die(.)ROsa noch einen <u>WUNSCHstern.</u>

Merle hatte bereits in der dritten Sequenz angemerkt, dass sie auch Alpträume habe (Z. 235-236). Die Erzieherin war mit einer Nachfrage kurz darauf eingegangen, weiter wurde aber nicht darüber gesprochen. Jetzt entwirft Merle selbst eine Lösung für ihr Problem: *ich frag MEIne eltern AUCh mal ob ich n TRAUMfänger haben kann.||* (Z. 279-280).

Der Erzieher wendet nun sich abschließend an Lilli (Z. 281-284): *na da BIN ich ja ma jetzt geSPANNT.| wenn die pauLine ihren traumfänger (.) MITnimmt,|ob DEIne träume dann SCHÖner werden.|| da BIN ich ja mal gespannt was DU erZÄHLST.||*. Er greift also Paulines Idee zur Erfüllung von Lillis Wunsch noch einmal auf und regt an, dass sie in der Realität umgesetzt wird. Darüber hinaus formuliert er seine innere Beteiligung am Fortgang des Themas (*na da BIN ich ja ma jetzt geSPANNT* Z. 281) und ermuntert Lilli, es

in einem der nächsten Gesprächskreise noch einmal anzusprechen. Mit *oKAY*.|| (Z. 285) beendet er die Behandlung von Lillis Wunschstern und leitet zu dem von Rosa über (Z. 286).

4.3. Formulierungen für inneres Erleben im Analysebeispiel

Im Folgenden richtet sich das Augenmerk, wie in anderen empirischen Studien (z. B. Aldridge, Wood 1997; Bretherton, Beeghly 1982; Shatz *et al.* 1983; Kauschke 2012a und 2012b; Klann-Delius 2012; Kristen *et al.* 2012) zum Erwerb des Vokabulars für innere Zustände und Prozesse, auf den Anteil an Formulierungen zu innerem Erleben im Analysebeispiel. Von Interesse ist dabei, welche Formulierungen vorkommen, wie häufig sie sind, von wem sie verwendet werden und welchen Inhaltskategorien sie sich zuordnen lassen.

4.3.1. Auftretenshäufigkeit

Die Gesamtwortzahl des analysierten Beispiels beträgt 491 Wörter. Den größeren Anteil im Beispiel produzieren die Kinder (285 Wörter = 58,04%). Der Wortanteil der Fachkräfte ist mit 206 Wörtern (41,96%) etwas geringer: Der Erzieher moderiert hauptsächlich die Bearbeitung von Lillis Anliegen (193 Wörter), die Erzieherin hält sich zurück und unterstützt ihn nur gelegentlich (13 Wörter).

Von den sieben anwesenden Kindern beteiligen sich fünf mit verbalen Äußerungen am Gespräch. Die meisten Redebeiträge liefern Pauline (135 Wörter), Lilli (64 Wörter) und Julia (62 Wörter): Lilli setzt mit ihrem Wunschstern das Thema, Pauline macht einen Lösungsvorschlag und Julia erzählt (wie auch Lilli und Pauline) über eigene Alpträumerfahrungen. Merle (18 Wörter) und Rosa (sechs Wörter) sprechen deutlich weniger: Merle berichtet über ihre Alpträumerfahrungen. Rosa kündigt zu Beginn einen Wunschstern an, der im Anschluss an den hier betrachteten Ausschnitt behandelt wird. Zwei Kinder (Tina und Edgar) hören ausschließlich zu, zeigen ihre innere Beteiligung aber nonverbal an: Wenn ein Kind spricht, beugen sie sich vor, schauen es an und drücken ihre Anteilnahme mit ernstem, mitunter betroffenem Gesicht, aber auch mit Lachen oder gespielterm Erschrecken aus.

Im analysierten Beispiel wurden insgesamt 40 verschiedene Formulierungen für Innerpsychisches identifiziert. 16 Formulierungen stammen ausschließlich von den Kindern, 15 ausschließlich von den Fachkräften. Neun Formulierungen werden sowohl von den Kindern als auch von den Fachkräften verwendet. Da die 40 Formulierungen im Analysebeispiel zum Teil mehrfach vorkommen, ergibt sich eine Gesamthäufigkeit von 92 Verwendungen, das sind 18,74% der Gesamtwortzahl. Auf die Kinder entfallen insgesamt 43

Verwendungen (15,09% ihrer Wortanzahl), auf die Fachkräfte 48 (23,30% ihrer Wortanzahl).

4.3.2. Inhaltskategorien

Bei der inhaltlichen Kategorisierung der sprachlichen Formulierungen für inneres Erleben sind wir bezogen auf den konkreten Kontext induktiv-deduktiv vorgegangen. Wir haben uns an vorhandenen Kategoriensystemen orientiert, die vor allem die Erwerbsreihenfolge des kindlichen Vokabulars für inneres Erleben fokussieren (z. B. Kauschke 2012a und 2012b; Klann-Delius 2012; Kristen *et al.* 2012). In der Tabelle 1 sind die innerpsychischen Bereiche und die dazugehörigen Kategorien mit den jeweils vorkommenden Formulierungen aufgeführt, ergänzt durch Häufigkeitsangaben bei den Kindern (erste Zahl nach der jeweiligen Formulierung) und den Fachkräften (zweite Zahl).

Emotion / Affekt	Volition / Fähigkeit	Bewertung / Graduierung	Kognition
*36 (**20 / 16)	13 (9 / 4)	37 (10 / 27)	6 (4 / 2)
Emotionswörter gruselig 2 / 0 gespannt 0 / 2	wollen 1 / 0 kann / kannst 4 / 4 könnte 4 / 0	Bestätigungsformeln okay 1 / 5 Adjektive, Adverbien nett 0 / 3 klasse 0 / 1 schön, schöner 1 / 1 Verben finde (ich nett) 0 / 3 Graduierungspartikeln nur 0 / 1 nur noch 0 / 1 mal 0 / 2 auch mal 1 / 0 wirklich 0 / 1 sehr 0 / 1 ganz 1 / 0 bisschen 1 / 0 ja 2 / 6 ja auch 2 / 0 eh 1 / 0 aber 0 / 1 gar nicht 0 / 1	weiß (du) 1 / 0 vorstellen 0 / 1 geträumt 3 / 1
Emotive Wörter schlimme Träume 3 / 2 was Schlimmes 1 / 0 Alpträume 2 / 1 schöne Träume 0 / 2 gestritten 1 / 0 vertrage 1 / 0			
Metaphern Wunschstern 1 / 3 Traum(weg)fänger 7 / 4 Biest 1 / 0			
Interjektionen oh 1 / 1 ei 0 / 1			

Tabelle 1
Häufigkeitsverteilung von Formulierungen des inneren Erlebens nach Inhaltskategorien

(* Verwendungen insgesamt; **Verwendungen der Kinder / der Fachkräfte).

Im Analysebeispiel wurden vier innerpsychische Bereiche identifiziert, die zum Teil mehrere Inhaltskategorien umfassen:

- **Emotion / Affekt** (13 Formulierungen, insgesamt 36 Verwendungen)
Darunter zählen *Emotionswörter* mit einem „direkten Bezug zu einer spezifischen Emotion“ (Kauschke 2012a, S. 6) sowie emotive und affektive Wörter, die „Gefühle ausdrücken, ohne diese explizit zu benennen“ bzw. „im Kern keinen emotionalen Bedeutungsgehalt haben, aber dennoch eine emotionale Wirkung auslösen können“ (ebd.). Außerdem fallen darunter *metaphorische Formulierungen* (so auch bei Christmann, Groeben 2011; Kauschke 2012a), also bildliche Ausdrücke mit übertragener Bedeutung (Vogt, Indefrey 2017), die dazu dienen können, „schwer fassbare, komplexe Gefühlszustände zu konkretisieren“ (Kauschke 2019, S. 263). Schließlich werden darunter *Interjektionen* erfasst (so auch ebd., S. 262), die ebenfalls emotionale Einstellungen ausdrücken, ohne sie direkt zu bezeichnen.
- **Volition / Fähigkeit** (3 Formulierungen, 13 Verwendungen)
Diesem Bereich werden verbale Ausdrücke zugeordnet, die eigene und fremde Absichten und Fähigkeiten zum Ausdruck bringen.
- **Bewertung / Graduierung** (21 Formulierungen, 37 Verwendungen)
Hier werden persönliche Bewertungen und moralische Urteile sowie deren Graduierungen aufgeführt. Darunter zählen *Bestätigungsformeln*, bewertende *Adjektive*, *Adverbien* und *Verben* sowie *Partikeln*. Allen ist gemeinsam, dass sie die subjektive Einstellung der sprechenden Person zum Gesagten ausdrücken und ggf. graduieren können.
- **Kognition** (3 Formulierungen, 6 Verwendungen)
Erfasst werden hier „mentale Wörter, meist Verben, die auf kognitive Zustände [und Prozesse, die Verf.] wie Denken, Glauben, Vorstellen, Vergessen verweisen“ (Kauschke 2012a, S. 7).

Weitere Bereiche des inneren Erlebens sind körperlich erfahrbare Zustände (z. B. Hunger oder Müdigkeit) sowie Wahrnehmungen (z. B. Geruchs- und Tastempfindungen) (vgl. Bretherton, Beeghly 1982; Kauschke 2012a und b; Kristen *et al.* 2012). Diese Bereiche sind im Analysebeispiel nicht vertreten. Im Folgenden werden die Befunde im Analysebeispiel erläutert, indem zu allen Inhaltskategorien die jeweiligen Formulierungen und ihre Verwendungen betrachtet werden.

Emotion / Affekt

Dieser Bereich ist im Analysebeispiel mit 36 Verwendungen am zweithäufigsten vertreten. Es finden sich zwei *Emotionswörter*: Das Adjektiv „gruselig“ wird zweimal von Lilli verwendet, dabei verweist sie

möglicherweise auf ihre Rezeptionserfahrungen mit vorgelesenen Gruselgeschichten. Das Emotionswort „gespannt“ wird vom Erzieher zweimal in demselben Turn verwendet, er drückt damit seine innere Beteiligung an den Äußerungen der Kinder aus.

Weiterhin finden sich sechs *emotive bzw. affektive Formulierungen*, die insgesamt 13mal verwendet werden, achtmal von den Kindern und fünfmal von den Fachkräften. Das Wort „Alpträume“ kommt dreimal vor. Eingeführt wird es vom Erzieher, Julia greift es in ihrer Second Story zweimal auf. Eng verwandt damit sind die Bezeichnungen „schlimme / schöne Träume“. Der Begriff „schlimme Träume“ wird von Lilli eingebracht und zweimal verwendet, auch der Erzieher greift ihn zweimal auf. Das Antonym „schöne Träume“ wird vom Erzieher erstmals verwendet und von der Erzieherin aufgegriffen. Die Verben „gestritten“ und „vertragen“ verwendet Pauline je einmal in ihrer Second Story über den Streit-Versöhnungs-Traum.

Im Analysebeispiel werden drei *metaphorische Formulierungen* 16mal verwendet (neunmal von den Kindern, siebenmal von den Fachkräften), sie bilden damit den Hauptanteil von Emotionswort-Verwendungen. Sie weisen einen starken emotionalen Gehalt auf. „Wunschstern“ (fünfmal verwendet) bildet zusammen mit „Stolperstein“ die zentrale Metapher des Gesprächskreisprojektes (s. o.). „Traum(weg)fänger“ ist die zentrale Metapher des Anliegens von Lilli und kommt im Ausschnitt neunmal vor. „Biest“ wird von Lilli einmal verwendet und verweist möglicherweise wie „gruselig“ auf ihre Rezeptionserfahrungen mit Grusel- und Verwünschungsgeschichten.

Weiterhin enthält das Analysebeispiel zwei *Interjektionen* mit emotionalem Gehalt: „Oh“ wird einmal von Lilli als verstärkender Kommentar zum Emotionswort „gruselig“ im fiktiven Kontext der Gruselgeschichte (Z. 278) verwendet und einmal vom Erzieher als Einleitung zur Würdigung von Paulines Angebot (Z. 222: *oh DAS ist ja NETT von dir.*||). „Ei“ wird einmal vom Erzieher verwendet (Z. 238: *ei dann MACHen wir das mal so;*||).

Volition / Fähigkeit

Dieser Bereich ist im Analysebeispiel durch die *Modalverben* „wollen“ und „können“ vertreten, die insgesamt 13mal verwendet werden (neunmal von den Kindern, viermal vom Erzieher). Das Modalverb „wollen“ benutzt Julia einmal (Z. 254: *und MAMA und PAPA wollen mir auch keinen KAUFen.*||). Das Modalverb „können“ im Indikativ wird viermal von den Kindern und viermal vom Erzieher verwendet, und zwar in der Form „(ich) kann / kann (ich)“ bzw. „kannst (du)“ in Verbindung mit verschiedenen Verben („gehen, sagen, mitnehmen, ausborgen, haben, vorstellen“). Der Erzieher verwendet es unter anderem in der exklamatorischen Äußerung *°hhh das KANN ich mir ja gar nicht !^VOR!stellen;*|| (Z. 268) in Verbindung mit einem syntaktischen Ausdruck von Emotionalität (Inversion).

Pauline nutzt das Modalverb „können“ für ihren Vorschlag, Lilli ihren Traumfänger mitzubringen, auch im Konjunktiv II. Innerhalb eines Turns kommt „könnte“ viermal vor: *ich KÖNNte ihn ja mal mit in n KINdergarten nehmen und dann KÖNNte ich ihr sie-|^h ihr solange sie keinen HAT,| ^h AUSborgen und wenn SIE dann einen^h HAT,| könnte sie mir den ja dann WIEdergeben.*|| (Z. 223-226). „Könnte“ wurde als eigenständige Formulierung gezählt, denn durch die Verwendung des Irrealis rahmt Pauline ihren Vorschlag als Möglichkeit, überlässt es aber Lilli, ihn anzunehmen oder nicht.

Bewertung / Graduierung

Dieser Bereich ist im Analysebeispiel mit insgesamt 21 Formulierungen in 37 Verwendungen am stärksten vertreten. Die pädagogischen Fachkräfte nutzen Ausdrücke aus diesem Bereich dreimal so häufig wie die Kinder (27mal versus zehnmal). Es findet sich die *Bestätigungsformel* „Okay“. Sie wird einmal von einem Kind (Julia in Z. 239) verwendet und fünfmal von den Fachkräften. Des Weiteren wurden vier bewertende *Adjektive und Adverbien* („nett“, „klasse“, „schön / schöner“) identifiziert, die fünfmal von den Fachkräften und einmal von einem Kind verwendet werden. Darüber hinaus kommt das *Verb* „finden“ in der bewertenden Äußerung „das finde ich (sehr nett)“ vor, es wird dreimal vom Erzieher verwendet. *Partikeln zur Abtönung bzw. Graduierung* des emotionalen Intensitätsgrades von Äußerungen kommen insgesamt 13mal vor. Sie werden 22mal verwendet und sind damit die im untersuchten Beispiel am häufigsten vertretene Bewertungskategorie. Die Fachkräfte setzen Partikeln öfter ein (14mal) als die Kinder (achtmal). Am häufigsten sind „ja“ (z. B. *^hhh das KANN ich mir ja gar nicht !^VOR!stellen;||* Z. 268) und „ja auch“ (z. B. *ich KANN ihr ja auch mal MEInen,| ... ausborgen* Z. 219) mit zehn Verwendungen (viermal bei den Kindern, sechsmal beim Erzieher).

Kognition

Dieser Bereich ist im Analysebeispiel mit sechs Verwendungen und drei Begriffen am seltensten vertreten. Es kommen die drei *Kognitionsverben* „wissen“, „vorstellen“ und „träumen“ vor. Einmal verwendet Julia „weißt du“ im Zusammenhang mit der Anrede des Erziehers. Die Formulierung „(Kann ich mir ja gar nicht) vorstellen“ wird einmal vom Erzieher verwendet, als er seine Verwunderung über Lillis Traum ausdrückt. Das Wort „geträumt“ wird insgesamt viermal verwendet (dreimal von den Kindern und einmal vom Erzieher).

5. Diskussion

Häufigkeit der Formulierungen für inneres Erleben

Insgesamt betrachtet ist im Analysebeispiel der Anteil an Ausdrücken zum innerpsychischen Erleben verhältnismäßig hoch (Kinder 15,09%; Fachkräfte 23,30% ihres Wortanteils), wenn man die Zahlen mit anderen Studien vergleicht, in denen allerdings deutlich jüngere Kinder untersucht wurden. Klann-Delius und Kauschke (1996) beobachteten bei 36 Monate alten Kindern einen Anteil von 6,6% bzw. 9% (Kauschke 2007). Der hohe Anteil an Ausdrücken für innerpsychisches Erleben im analysierten Beispiel könnte mit dem Thema des behandelten Anliegens zu tun haben. Es geht um Alpträume, die in der Regel emotional stark belastend sind, und um Lillis Wunsch, ihre schlimmen Träume mit einem Traumfänger zu bekämpfen.

Inhaltskategorien

Die beiden Bereiche *Emotion / Affekt* und *Bewertung / Graduierung* kommen deutlich häufiger vor als *Volition / Fähigkeit* und *Kognition*. Das ist nachvollziehbar, wenn man das behandelte Thema betrachtet: Es geht um die Bearbeitung eines Wunsches (Traumfänger bekommen), um ein emotional belastendes Problem loszuwerden (Alpträume haben). Dieses Anliegen begründet die hohe Anzahl an *emotiven und affektiven Wörtern* (vgl. Tabelle 1). Des Weiteren gebrauchen die Kinder *Metaphern*, vor allem bei der Kategorisierung eines Anliegens als Wunsch („Wunschstern“) bzw. als Störung / Problem („Stolperstein“). Das stellt hohe kognitive Anforderungen an sie und oft helfen die Fachkräfte dabei (ausführlich dazu Bose, Kurtenbach 2019). Im Verlauf des Projekts werden die Kinder immer sicherer darin, Themen passend als „Wunschstern“ bzw. „Stolperstein“ zu kategorisieren. Das führt sogar dazu, dass sie beide Begriffe auf den Kindergarten-Alltag übertragen und Ereignisse metasprachlich in diesen Kategorien reflektieren (ebd.). Die hohe Anzahl an *bewertenden und graduierenden Wörtern* kommt dadurch zustande, dass sowohl die Kinder als auch die pädagogischen Fachkräfte mit zahlreichen Partikeln („nur noch“, „wirklich“, „sehr“ usw.) ihren Äußerungen einen emotional-bewertenden Gehalt verleihen. Außerdem geht der Erzieher mit einer Vielzahl von bestätigenden Äußerungen auf die Anliegen und Lösungsvorschläge der Kinder ein.

Dass Formulierungen aus dem Bereich der *Kognition* seltener verwendet werden als solche aus den anderen innerpsychischen Bereichen, könnte mit der Erwerbsreihenfolge im Zusammenhang stehen. Bretherton und Beeghly (1982), Jenkins *et al.* (2003) und Kristen *et al.* (2014) weisen zum Beispiel aus, dass Kinder in diesem Bereich die wenigsten Begriffe beherrschen. Es scheint sich also um einen Bereich zu handeln, welcher erst mit zunehmendem Alter entwickelt wird. Formulierungen für *körperlich erfahrbare Zustände* sowie

Wahrnehmungen sind im Analysebeispiel nicht vertreten. Das lässt sich mit dem Thema des Gespräches erklären, in welchem solche Inhalte nicht thematisiert werden.

Interaktiver Umgang

In der Gesamtschau zeigt sich, dass viele Bezeichnungen für inneres Erleben mehrmals vorkommen. Das ist auf die Gesprächsdynamik der Themenbehandlung zurückzuführen: Die Kinder und Fachkräfte beziehen sich in ihren Äußerungen aufeinander (z. B. Z. 203-204; 208-209; 234-235; 260-261). Sie greifen Begriffe und Formulierungen, die sie selbst oder die anderen Gesprächsbeteiligten bereits verwendet haben, wieder auf (z. B. „Traum(weg)fänger“, „schlimme Träume“). Gelegentlich nuancieren sie deren Bedeutung, bringen Synonyme oder Antonyme ein (z. B. „schlimme Träume“ – „schöne Träume“ – „Alpträume“). Die Gruppe bleibt so über mehrere Turns bei einem Anliegen und entwickelt es gedanklich weiter (ausführlich dazu Bose *et al.* 2020; Wala *et al.* under review): Das Anliegen wird kategorisiert und spezifiziert, die Gruppe sucht nach Lösungsideen und wägt deren jeweilige Passung ab. Diese themenzentrierte Gesprächsarbeit ist typisch für das Gesamtkorpus. Sie wird dadurch möglich, dass es ausschließlich um Themen geht, an denen den Kindern etwas liegt, in die sie auch häufig stark emotional involviert sind. Im Korpus kommt es mehrfach vor, dass Anliegen über mehrere Gesprächskreise hinweg erneut aufgegriffen und bearbeitet werden. So wird zum Beispiel auch das Thema „Traumfänger“ noch einmal kurz im darauf folgenden Gesprächskreis behandelt: Pauline erzählt, dass sie vergessen habe, ihren Traumfänger mitzubringen, und der Erzieher berichtet der etwas anders zusammengesetzten Gruppe über Lillis Wunschstern aus dem letzten Gesprächskreis.

Indem die Fachkräfte den Lösungsvorschlägen und Erzählungen der Kinder Raum geben, ermuntern sie sie, Ideen für die Bearbeitung ihrer Anliegen zu entwickeln. Durch das Wiederholen von Formulierungen können sich Begriffe etablieren und von den Kindern erworben werden.

Gesprächsförderliches Potential

Wie der hauptsächlich moderierende Erzieher sein eigenes inneres Erleben und das der Kinder verbal, aber auch paraverbal ausdrückt, ist exemplarisch für das Agieren beider Fachkräfte in zahlreichen anderen Themenbehandlungen. Der Erzieher vermittelt beständig seine emotionale Beteiligung, sowohl bei der Verständnissicherung während der Formulierung und Spezifizierung von Lillis Anliegen als auch bei der wohlwollenden Bestätigung der kindlichen Lösungsvorschläge und Schilderungen eigener Erfahrungen (s. 4.2. und 4.3.2.). Durch die Formulierungshilfen, Bestätigungen und Nachfragen der pädagogischen Fachkräfte können die Kinder modellhaft lernen, ihrem innerpsychischen Erleben Ausdruck zu verleihen. Dass solche Formulierungen

aufgegriffen und abgewandelt werden, zeigt sich im analysierten Beispiel vor allem dadurch, dass von den insgesamt 40 verschiedenen Formulierungen für inneres Erleben 18 mehrfach vorkommen. Empirisch belegt worden ist bereits, dass der Sprachgebrauch von engen Bezugspersonen (z. B. Müttern und Geschwistern) im Hinblick auf die Verwendung von Begriffen für innere Zustände und Prozesse einen positiven Einfluss auf die Entwicklung des Emotionsvokabulars und -verständnisses von Kindern hat (z. B. Bretherton, Beeghly 1982; Jenkins *et al.* 2003; Kauschke 2007). Es ist anzunehmen, dass dies auch für den Einfluss pädagogischer Fachkräfte auf die Sprachenwicklung von Kindern gilt.

6. Fazit und Ausblick

Grundsätzlich verfolgt das Gesprächskreisprojekt „Stolpersteine und Wunschsterne“ das Ziel, Wünsche und Sorgen der Kinder zum Thema zu machen, um diese in der Gruppe interaktiv zu bearbeiten und nach Lösungsideen zu suchen. Im Beitrag wurde untersucht, welche Ausdrucksmittel die Beteiligten nutzen, um miteinander über Inneres zu sprechen, und wie sie interaktiv damit umgehen. Im Mittelpunkt stand die Behandlung eines kindlichen Anliegens, aber es wurde auch Bezug auf das Gesamtkorpus 2015/16 genommen.

Es zeigt sich, dass die Fachkräfte und die Kinder über eine Vielzahl an verbalen, paraverbalen und nonverbalen Ressourcen für den Ausdruck inneren Erlebens verfügen. Im Analysebeispiel nutzen sie verhältnismäßig viele verbale Formulierungen, insbesondere aus den Bereichen *Emotion / Affekt* sowie *Bewertung / Graduierung*. Im Verlauf der Anliegensbehandlung wird eine Vielzahl von Formulierungen des inneren Erlebens mehrfach verwendet und dabei auch modifiziert – sowohl von den Kindern als auch von den pädagogischen Fachkräften. Durch eine solche empathische Gesprächsdynamik können Begriffe und Formulierungen etabliert und von den Kindern kontextpassend erworben werden.

Innere Beteiligung kommt aber auch paraverbal zum Ausdruck. So zeigt der Erzieher seine Anteilnahme an den Anliegen und Ideen der Kinder mit einem durchgehend freundlichen Sprechausdruck an (gekennzeichnet z. B. durch legatohafte Melodiebewegungen, hellen Stimmklang, melodisch-emphatische Akzente). Die Kinder verwenden unterschiedliche Sprechausdruckformen: Wenn sie besonders ernste Themen einbringen (wie im analysierten Beispiel), sprechen sie zumeist in einem reduzierten Melodieumfang (monoton), eher langsam und leise. Es finden sich zahlreiche weitere Stellen im Gesamtkorpus, in denen solche Sprechausdrucksformen vorkommen, wenn die Kinder ernste, emotional belastende Themen ansprechen. Im Verlauf der analysierten Anliegensbehandlung verändert sich

der Sprecherausdruck, insbesondere dann, wenn die Kinder in eine fiktive Modalität wechseln. So malt Lilli bei der Schilderung ihres fiktiven Traums den Grusel vor dem Biest, in das sich ihre Freundin Mara verwandelt hat, prosodisch expressiv aus (lautes Sprechen mit großen Melodiebewegungen, starken Lautdehnungen und empathischen Akzenten).

Nonverbal zeigen die Kinder ihre innere Anteilnahme, indem sie einander bei ernsten Themen mit ernsten Gesichtern anschauen, sich körperlich der sprechenden Person zuwenden oder einander bei den Lösungsideen aufmunternd zulächeln. Mehrfach ist im Korpus ein maskierter Gesichtsausdruck zu beobachten, zum Beispiel wenn die Kinder (wie im analysierten Ausschnitt) ernste Anliegen mit einem starren, unsicher-verlegenen Lächeln vorbringen. Ob diese Maskierung dadurch bedingt ist, dass manche Themen bereits schambesetzt sind, oder ob möglicherweise eine gewisse Hemmung, vor der ganzen Gruppe zu sprechen, die Ursache ist, konnte bis jetzt nicht geklärt werden.

Für das ausgewählte Anliegen (Lillis Wunschstern) wurde einerseits gesprächsanalytisch rekonstruiert, welche Ressourcen die beteiligten Kinder und pädagogischen Fachkräfte beim Sprechen über inneres Erleben nutzen und wie sie interaktiv damit umgehen. Andererseits wurden die verwendeten Formulierungen quantitativ betrachtet und inhaltlich kategorisiert. Dabei zeigte sich, dass eine klare und eindeutige Kategorisierung von Begriffen des inneren Erlebens nicht immer möglich ist. Bei der Zuordnung konnte der konkrete Gesprächskontext eine Orientierung liefern. Damit ergibt sich aus der Verknüpfung von qualitativem und quantitativem Vorgehen ein Mehrwert. Denn so können Ausdrucksformen des inneren Erlebens im konkreten Gesprächskontext erklärt und miteinander in Beziehung gesetzt werden.

Das Analysebeispiel stammt aus einem umfangreichen Korpus von 16 Gesprächskreisen (Periode 2015/16), in denen sechsjährige Kinder mit den pädagogischen Fachkräften über 10 Monate hinweg insgesamt 109 Anliegen verhandeln. Darüber hinaus liegt ein Videokorpus von Gesprächskreisen mit fünfjährigen Kindern vor (Periode 2018/19), so dass Vergleichsbefunde erhoben werden können. Anhand beider Korpora soll zukünftig untersucht werden, in welchen Sequenzen gehäuft inneres Erleben angesprochen wird, ob es im Verlauf einer Periode Veränderungen bzw. Entwicklungen im kindlichen und erwachsenen Gebrauch von Ausdrucksmitteln für inneres Erleben gibt, ob bestimmte Formulierungen besonders typisch für die Kinder oder die Erwachsenen sind, ob und ab wann bestimmte Ausdrücke von anderen Beteiligten dauerhaft übernommen werden. So können die in der vorliegenden Analyse exemplarisch gewonnenen Erkenntnisse auf der Basis größerer Datenmengen anhand der oben erwähnten Datenbank überprüft und generalisierbare Ergebnisse erzielt werden.

Bionote: Ines Bose ist Diplom-Sprechwissenschaftlerin und seit 2009 außerplanmäßige Professorin an der Martin-Luther-Universität Halle-Wittenberg. 1989 wurde sie in Sprechwissenschaft / Phonetik promoviert mit einer Arbeit zur phonetischen Analyse frei formulierter Gesprächsäußerungen und habilitierte sich 2003 mit einer Arbeit zur empirischen Analyse des Sprechausdrucks in kindlichen Rollenspielen. Aktuell forscht und lehrt sie unter anderem zur Phonetik und Rhetorik der Gesprächsfähigkeit, insbesondere zur Entwicklung der kindlichen Kommunikationsfähigkeit (Sprechausdruck, Argumentation). Dr. Stephanie Kurtenbach ist Klinische Sprechwissenschaftlerin. Nach dem Studium der Sprechwissenschaft (Diplom 1999) hat als Sprachtherapeutin gearbeitet. Seit 2006 lehrt sie an der Abteilung Sprechwissenschaft und Phonetik der Martin-Luther-Universität in Halle (Promotion 2008). Ihre Schwerpunkte in Forschung, Lehre und Therapie sind: Frühdiagnose und Intervention bei sprachauffälligen Kindern, Spracherwerbsstörungen bei sensorisch-integrativen Dysfunktionen, Prävention (Elternarbeit, Arbeit mit pädagogischen Fachkräften). Seit mehreren Jahren leitet sie Kooperationsprojekte mit pädagogischen Einrichtungen zu den Themen Sprachbildung und -förderung. Zusatzqualifikationen: Sensorische Integrationstherapeutin, systemische Beraterin, Prompt-Therapeutin Level 1, RIT-Trainerin.

Adressen der Autorinnen: ines.bose@sprechwiss.uni-halle.de;
stephanie.kurtenbach@sprechwiss.uni-halle.de

Literatur

- Aldridge M. and Wood J. 1997, *Talking about Feelings: Young Children's Ability to express emotions*, in "Child Abuse & Neglect" 21 [12], pp. 1221-1233.
- Bahn D., Vesker M., Garcia Alanis J.C., Schwarzer G. and Kauschke Ch. 2017, *Age-Dependent Positivity-Bias in children's Processing of Emotion Terms*, in "Frontiers in Psychology" 8, 1268, pp. 1-16. <https://doi.org/10.3389/fpsyg.2017.01268>.
- Baron-Cohen S., Golan O., Wheelwright S., Granader Y. and Hill J. 2010, *Emotion word comprehension from 4 to 16 years old: a developmental survey*, in "Frontiers in Evolutionary Neuroscience" 109 [2], pp. 1-8. <https://doi.org/10.3389/fnevo.2010.00109>.
- Bose I., Hannken-Illjes K. und Kurtenbach S. 2020, *und was MACHen wir da jetzt um das problem zu LÖsen? || Argumentation im Kindergarten und ihre Förderung in einem Gesprächskreisprojekt einer Kindertagesstätte*, in „Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion" 21, S. 1-29. www.gespraechsforschung-ozs.de (15.05.2023).
- Bose I. und Kurtenbach S. 2019, „*Stolpersteine und Wunschsterne*“ – Förderung von Partizipation und Gesprächsfähigkeit in der Kita, in Bose I., Hannken-Illjes K. und Kurtenbach S. (Hg.), *Kinder im Gespräch*, Frank&Timme, Berlin u. a., S. 113-148.
- Bretherton I. and Beeghly M. 1982, *Talking about internal states: The acquisition of an explicit theory of mind*, in "Developmental Psychology" 18 [6], pp. 906-921.
- Bretherton I., Fritz J., Zahn-Waxler C. and Ridgeway D. 1986, *Learning to Talk about Emotions*, in "A Functionalist Perspective. Child Development" 57, pp. 529-548.
- Christmann U. und Groeben N. 2011, *Die Herstellung von Intersubjektivität bei figurativer Sprache*, in "Rhetorik" 30, S. 1-16. <https://doi.org/10.1515/9783110236385.1>.
- Deppermann A. 2008, *Gespräche analysieren. Eine Einführung*, VS Verlag für Sozialwissenschaften, Wiesbaden. <https://doi.org/10.1007/978-3-531-91973-7>.
- Grosse G., Streubel B., Gunzenhauser C. and Saalbach H. 2021, *Let's Talk About Emotions: the Development of Children's Emotion Vocabulary from 4 to 11 Years of Age*, in "Affective Science" 2, pp. 150-162. <https://doi.org/10.1007/s42761-021-00040-2>.
- Hildebrandt F., Scheidt A., Hildebrandt A., Hédervári-Heller É. und Dreier A. 2016, *Sustained shared thinking als Interaktionsformat und das Sprachverhalten von Kindern*, in „Frühe Bildung" 5 [2], Hogrefe, Göttingen, S. 82-90.
- Jenkins J.M., Turell S.L., Kogushi Y., Lollis S. and Ross H.S. 2003, *A longitudinal investigation of the dynamics of mental state talk in families*, in "Child Development" 74 [3], pp. 905-920.
- Kauschke Ch. 2012a, *Sprechen über Inneres - die Versprachlichung von Emotionen im Kindesalter*, SAL-Bulletin 145, S. 5-16.
- Kauschke Ch. 2012b, *Kindlicher Spracherwerb im Deutschen. Verläufe, Forschungsmethoden, Erklärungsansätze*, de Gruyter, Berlin.
- Kauschke, Ch. 2007, *Sprache im Spannungsfeld von Erbe und Umwelt*, in „Die Sprachheilarbeit" 52, S. 4-16.
- Kauschke Ch. 2019, *Linguistische Perspektiven auf Emotion und Sprache*, in Kappelhoff H., Bakels J.-H., Lehmann H. und Schmitt Ch. (Hg.), *Emotionen – Ein interdisziplinäres Handbuch*, Metzler © Springer, Heidelberg, S. 262-270.
- Klann-Delius G. 2012, *Spracherwerb. Eine Einführung*, J.B. Metzler, Stuttgart.
- Klann-Delius G. 2015, *Emotion in Language*, in Lüdtkke U.M. (Hg.), *Emotion in Language. Theory – Research – Application*, John Benjamins, Philadelphia, pp. 135-156.

- Klann-Delius G. und Kauschke Ch. 1996, *Die Entwicklung der Verbalisierungshäufigkeit von inneren Zuständen und emotionalen Ereignissen in der frühen Kindheit in Abhängigkeit von Alter und Affekttyp: eine explorative, deskriptive Längsschnittstudie*, in „Linguistische Berichte“ 161, S. 68-89.
- Kristen S., Sodian B., Licata M., Thoermer C. and Poulin-Dubois D. 2012, *The Development of Internal State Language during the Third Year of Life: A Longitudinal Parent Report Study*, in “Infant and Child Development” 21, pp. 634-645. <https://doi.org/10.1002/icd.1767>.
- Kristen S., Chiarella S., Sodian B., Aureli T., Genco M. and Poulin-Dubois D. 2014, *Crosslinguistic Developmental Consistency in the Composition of Toddlers' Internal State Vocabulary: Evidence from Four Languages*, in “Child Development Research” 8, 575152, pp. 1-8. <http://dx.doi.org/10.1155/2014/575142>.
- Kurtenbach S., Bose I. und Hannken-Illjes K. 2019, *Argumentative Fähigkeiten im Vorschulalter – eine korpusbasierte Analyse*, in „Forschung Sprache. E-Journal für Sprachheilpädagogik, Sprachtherapie und Sprachförderung“ 2, S. 26-36. <https://www.forschung-sprache.eu/heft-archiv/jahrgang-7-2019/heft-2/> (15.05.2023).
- Nelson K. 1996, *Language in Cognitive Development. Emergence of the Mediated Mind*, Cambridge University Press, Cambridge.
- Petermann F. und Wiedebusch S. 2016, *Emotionale Kompetenz bei Kindern*, Hogrefe, Göttingen u. a.
- Röska-Hardy L. 2011, *Der Erwerb der Theory-of-Mind-Fähigkeit. Entwicklung, Interaktion und Sprache*, in Leimbrink K., Quasthoff U., Hoffmann L. (Hg.), *Die Matrix der menschlichen Entwicklung*, De Gruyter, Berlin, S. 96-142.
- Selting M., Auer P., Barth-Weingarten D., Bergmann J., Bergmann P., Birkner K., Couper-Kuhlen E., Deppermann A., Gilles P., Günthner S., Hartung M., Kern F., Mertzlufft Ch., Meyer Ch., Morek M., Oberzaucher F., Peters J., Quasthoff U., Schütte W., Stukenbrock A. und Uhmann S. 2009, *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)*, in „Gesprächsforschung“ 10, S. 353-402. <http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf> (15.05.2023).
- Shatz, M., Wellman, H. M. and Silber, S. 1983, *The acquisition of mental verbs: A systematic investigation of the first reference to mental state*. In: “Cognition” 14 [3], pp. 301-321.
- Siraj-Blatchford I. 2009, *Conceptualizing progression in the pedagogy of play and sustained shared thinking in early childhood education: a Vygotskian perspective*, in “Education and Child Psychology” 26 [2], pp. 77-89.
- Streubel B., Gunzenhauser C., Grosse G. and Saalbach H. 2020, *Emotion-specific vocabulary and its contribution to emotion understanding in 4- to 9-year-old children*, in “Journal of Experimental Child Psychology” 193, 104790. <https://doi.org/10.1016/j.jecp.2019.104790>.
- Vogt S. und Indefrey P. 2017, *Metaphernerwerb: eine empirische Studie bei Kindern im Alter von sechs bis vierzehn Jahren*, in Döring M., Jäkel O., Mutz K., Osthus D., Polzin-Haumann C. und Visser J. (Hg.), in „metaphorik.de“ 27. <https://www.metaphorik.de/de> (15.05.2023).
- Wala J., Hannken-Illjes K., Bose I. and Kurtenbach S. (under review), *Same here: Second stories and their argumentative function in child talk*, in “European Journal of Psychology of Education“. <https://www.springer.com/journal/10212>.
- Widen S.C. and Russell J.A. 2008, *Children acquire emotion categories gradually*, in “Cognitive Development” 23 [2], pp. 291-312.

Anhang: Legende der Transkriptionszeichen

(Feintranskript orientiert an GAT2; vgl. Selting *et al.* 2009)

[] []	Überlappungen und Simultansprechen
°h °hh °hhh	Einatmen, je nach Dauer (0,2-0,5; 0,5-0,8; 0,8-1,0 s)
h° hh° hh°	Ausatmen, je nach Dauer
(.)	Mikropause
(-) (--) (---)	kurze, mittlere, längere Pausen (0.25 - 0.75 s bis ca. 1 s)
: :: :::	Lautdehnung, je nach Dauer (0,2-0,5; 0,5-0,8; 0,8-1,0 s)
mh	einsilbiges Rezeptionssignal
mhm	zweisilbiges Rezeptionssignal
geSPANNT	akzentuierte Silbe
ge! SPANNT!	extrastark akzentuierte Silbe
˘ ˘ ˘	interlineare Akzenttonhöhenbewegung (steigend, fallend, steigend-fallend)
↑	Tomhöhen sprung zum Gipfel der Akzentsilbe
? , - i .	Tonhöhenbewegung am Ende von prosodischen Einheiten (hoch steigend, mittel steigend, gleichbleibend, mittel fallend, stark fallend)
	Abgeschlossenheitsgrad der Sprechereinheit (inhaltlich-prosodisch abgeschlossen, nicht abgeschlossen)
<<len>>	Sprechgeschwindigkeitsveränderung (lento) mit Reichweite
<<mf f ff p pp>>	Lautstärkeveränderung (mezzoforte, forte, fortissimo, piano, pianissimo) mit Reichweite
<<mon ruf :-)>>	Sprechausdrucksmodifikation (monoton, gerufen, lächelnd) mit Reichweite lächelnd gesprochen
<<mon>>	monoton gesprochen
<<ruf>>	gerufen
<<"erschrocken">>	gespielter Sprechausdruck
((nickt))	nonverbale Handlung mit Reichweite
<<@>>	gerichtet an (Name) mit Reichweite

DIE DATENBANK „STUDENTISCHE REFERATE IN DER FREMDSPRACHE DEUTSCH“ UND IHRE ANWENDUNG IN EINER PILOTSTUDIE ZUR RHETORISCHEN SPRECHWIRKUNG

ULRIKE A. KAUNZNER, LISA BARDECK
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Abstract – This article introduces a database designed to evaluate students’ public speaking proficiency at the University of Modena and Reggio Emilia: “La comunicazione accademica studentesca orale in LS (tedesco), L1 italiano” (Academic student oral communication in FL German, L1 Italian). Central to the project is the collection of audio and video recordings capturing student presentations delivered in German L2, which will be inserted into the BAS CLARIN Repository (Bavarian Archive for Speech Signals, LMU Munich). The corpus consists of audio and video material and is currently being constructed through various transcription phases, with special emphasis on the tool OCTRA for orthographic transcription and lexical recognition. A longitudinal study utilizes a sub-corpus of paired video recordings of the same students, taken two years apart, revealing valuable insights into rhetorical speech effects. The analysis aims to track changes or improvements in speaking proficiency over time, highlighting the development or persistence of skills, and exploring differences between self-assessment and external evaluations. Additionally, the study will investigate differences in presentation styles and effectiveness between in-person and online settings. Key aspects under examination include overall impression, clarity, naturalness, perceived competence, and speaking style, based on phonetic features such as speed, pausing, articulation precision, variation, and emphasis.

Keywords: rhetorical speech effect; self-assessment vs. external assessment; longitudinal study; transcription; audio/video corpus; German L2 database; speech analysis software.

*If I went back to college again,
I'd concentrate on two areas: learning
to write and to speak before an audience.
Nothing in life is more important than
the ability to communicate effectively.*
(Gerald R. Ford, 38. US-Präsident)

1. Einleitung

Die Fähigkeit, sich im akademischen Diskurs klar und überzeugend auszudrücken, rückt immer mehr ins Zentrum des universitären DaF-Unterrichts. Kompetent vorzutragen und zu präsentieren sind neben anderen rhetorischen Fähigkeiten und Fertigkeiten Schlüsselkompetenzen nicht nur für

den akademischen Wissenserwerb, sondern grundlegende Voraussetzungen für das spätere Berufsleben. In diesem Beitrag wird der Fokus auf die Bedeutung diverser Elemente gelegt, die bei der Redekompetenz eine entscheidende Rolle spielen und dazu beitragen, die inhaltliche Substanz verständlich und überzeugend zu vermitteln, wobei es im Grunde um die Frage geht, wie speziell DaF-Lernende diese Fähigkeiten und Fertigkeiten optimieren können.

Zu diesem Zweck wurde die Datenbank „Studentische Referate in der FS Deutsch“ an der italienischen Universität Modena und Reggio Emilia (Unimore) erstellt, welche öffentlich zugänglich¹ sein wird und somit zukünftig eine Quelle für diesbezügliche Fragestellungen darstellt. Die in diesem Artikel vorgestellte Longitudinalstudie verwendet ein Subkorpus mit im Abstand von zwei Jahren aufgenommenen Videoaufnahmen derselben Studierenden. Anhand von Fragebögen erfolgt ein Vergleich zwischen Selbst- und Fremdwahrnehmung, um nicht zuletzt die Langzeitwirkung universitärer Redeschulung zu überprüfen.

Der vorliegende Beitrag gliedert sich in vier Abschnitte:² Nach einem einführenden Kapitel wird der Fokus auf die Didaktik von Rederhetorik und die Bedeutung rhetorischer Kompetenzen bei DaF-Lernenden gelegt (Kapitel 2). Am Ende jedes Abschnitts wird auf die Relevanz in Bezug auf die Pilotstudie hingewiesen. Die Betrachtung erstreckt sich über den akademischen Diskurs bis zum Anspruch auf Verständlichkeit, wobei die Sprechwirkung nicht muttersprachlicher Sprecher im Vordergrund steht. Am Beispiel der humanistischen Studiengänge LCE (*Lingue e culture europee*) und LACOM (*Languages for communication in international enterprises and organizations*)³ an der Universität Modena und Reggio Emilia wird die Vermittlung rhetorischer Kompetenzen der an der Studie beteiligten Kurse zur deutschen Sprache und Sprachwissenschaft beleuchtet.

Kapitel 3 behandelt die Entstehung und Beschreibung der bei der Umfrage und Analyse zugrundeliegenden Datenbank. Ein Einblick in die Schaffung dieses Datenkorpus, bestehend aus Audio- und Videoaufnahmen von studentischen Referaten der oben erwähnten Kurse (Deutsch als Fremdsprache), wird gegeben. Die Einbindung dieses Korpus in die Datenbank BAS CLARIN Repository (Bavarian Archive for Speech Signals, LMU

¹ Vgl. <https://clarin.phonetik.uni-muenchen.de/BASRepository/> [10.06.2024].

² Die Abschnitte 3.2. (Datensammlung und Aufbereitung), 3.3. (Beschreibung der Software-Programme) und 4.2. (Methode: Online-Umfrage) wurden von Lisa Bardeck verfasst, die bei der Auswertung der Daten als externe Mitarbeiterin beteiligt war. Die Ergebnisse wurden gemeinsam formuliert.

³ Es handelt sich hier um ein Masterstudium, in dem sämtliche Kurse auf Englisch abgehalten werden, ausschließlich der spezifischen Kurse, die eine andere Sprache (im vorliegenden Fall die deutsche Sprache) betreffen. Vgl. LACOM Unimore, <https://www.unimore.it/it/didattica/corsi-di-studio/languages-communication-international-enterprises-and-organizations> [11.08.2024].

München)⁴ mit Fokus auf die im Projekt angewandte Transkription unter Einbeziehung des Tools OCTRA⁵, bildet den Schwerpunkt dieses Teils. Kapitel 4 stellt die empirische Studie zur rhetorischen Sprechwirkung studentischer Referate anhand grundlegender Erkenntnisse vor und veranschaulicht die Anwendung der in Kapitel 3 erläuterten Datenbank. Kapitel 5 widmet sich der Diskussion ausgewählter Ergebnisse und der Formulierung weiterführender Fragestellungen.

2. Rhetorische Kompetenzen bei DaF-Lernenden

Die Rhetorik, als abendländische Wissenschaftstradition im schulischen und universitären Kontext, ist bis in die Antike zurück zu verfolgen, wo die Kunst der überzeugenden Rede als essenzielle Fähigkeit für politische, juristische und soziale Interaktionen galt. Die rhetorischen Theorien von Sokrates (470–399 v. Chr.), Aristoteles (384–322 v. Chr.), Cicero (106–42 v. Chr.) und Quintilian (36-90 n. Chr.) prägten die humanistische Bildung des Mittelalters und der Renaissance und gelten in ihren Grundzügen bis heute (Ueding 2011; Erler, Tornau 2019; Fix *et al.* 2008). Viele der Lehren des ersten staatlich finanzierten römischen Rhetoriklehrers Quintilian (*Institutio Oratoria*) spiegeln sich auch in jüngeren Lehrwerken zu Rede- und Gesprächsrhetorik wider.

Während sich die antike Rhetorik stark auf persuasive Kommunikation gestützt hat, liegt der Sprechwissenschaft ein weiter gefasster Rhetorikbegriff zugrunde, indem sie von mündlicher Kommunikation spricht und dabei auch die ästhetische Kommunikation (Sprechkunst) und die therapeutische Kommunikation (Stimm-, Sprech-, Sprachtherapie) mit einbezieht.⁶

Was verstehen wir heute unter rhetorischen Kompetenzen im schulischen und universitären Kontext? Im sprechwissenschaftlichen Diskurs wird bei der Didaktik der mündlichen Kommunikation zwischen Gespräch und Rede unterschieden, konsequenterweise bei der rhetorischen Kommunikation zwischen Gesprächs- und Rederhetorik:

Der Terminus „rhetorische Kommunikation“ wurde gewählt und begründet als Oberbegriff für alle Erscheinungsformen mündlicher Rhetorizität. Er bezeichnet nicht nur in traditioneller Sicht „Formen der Rede“, sondern auch „Formen des Gesprächs“. Es ist also angebracht „Gesprächsrhetorik“ von „Rederhetorik“ zu unterscheiden, beide aber unter dem Oberbegriff Rhetorische Kommunikation zu subsumieren. (Geißner 2000, S. 126)

⁴ Für nähere Informationen über CLARIN Repository wird auf die Webseite <https://www.clarin.eu/> sowie Fišer, Witt (2022) verwiesen.

⁵ Der Name OCTRA steht für OrthographiC TRAnscription. <https://clarin.phonetik.uni-muenchen.de/apps/octra/octra/login> [24.07.2024].

⁶ Zum Rhetorikbegriff in der Sprechwissenschaft siehe u.a. Till (2015).

Bei beiden Formen gehe es nach Helmut Geißner um die „Reziprokhandlungen von Sprechdenken und Hörverstehen“ (1982, S. 98), im Falle der Rederhetorik folglich um „Redefähigkeit und Redeverstehensfähigkeit“ (ebd.). Das von beiden Formen verfolgte Ziel einer Handlungsauslösung könne nach Geißner mental („mitdenken“) oder real („mitwirken“, „mitarbeiten“) sein. Die Sprecherziehung als angewandte Sprechwissenschaft spricht folglich von Angewandter Rhetorik.

Heute bilden Seminare zur Angewandten Rhetorik (Rede- und Gesprächsrhetorik) in der beruflichen Aus- und Fortbildung in den unterschiedlichsten Sektoren, darunter Bankwesen und Marketing, einen festen Bestandteil des Kursangebots. Wie das Curriculum an der Unimore zeigt, ist auch im universitären DaF-Kontext ein Umdenken erkennbar. Schließlich umfassen sogar die allgemeinen Richtlinien zum Fremdsprachunterricht den Erwerb rhetorischer Kompetenzen (vgl. *Gemeinsamer europäischer Referenzrahmen für Sprachen*, Trim *et al.* 2001⁷). Es ist Anliegen der Sprecherziehung, beide Formen der Angewandten Rhetorik im DaF-Kontext zu vermitteln, dabei die kulturellen Elemente zwischenmenschlicher Kommunikation mit einzubeziehen und die rhetorischen Kompetenzen als Ziel der universitären Vermittlung entsprechend zu berücksichtigen. Leider sieht die Praxis oft noch anders aus, und im DaF-Unterricht stehen nach wie vor Grammatik und Vokabellernen im Vordergrund; auch beim Halten von Referaten wird dem Vortrag zugunsten inhaltlicher Kriterien verhältnismäßig wenig Aufmerksamkeit geschenkt. Das liegt in erster Linie an der mangelnden Zeit und an großen Gruppen, oft aber auch an der Ausbildung der Lehrkräfte, die nur selten eine entsprechende stimmlich-sprecherische und rhetorische Schulung genossen haben.⁸

Auch wenn die Sprecherziehung im DaF/DaZ-Studium oder auch generell im Lehramtsstudium nicht überall zum Pflichtprogramm zählt, oftmals gar nicht angeboten wird, so werden in Deutschland, aber auch in anderen Ländern, in jüngerer Zeit vermehrt Rufe laut, sie einzuführen.⁹ Im Ausland tätige Sprecherzieherinnen und Sprecherzieher können einen wichtigen Beitrag zu diesem Desiderat leisten. Angebote in der Gesprächs- und Rederhetorik werden in anderen Ländern als „besonders attraktiv“ (Forster 2004, S. 325) wahrgenommen, da sie oftmals als neu empfunden werden und eine Lücke im curricularen Angebot füllen. Das hier vorgestellte universitäre Projekt an der Universität Modena und Reggio Emilia zielt darauf ab,

⁷ Hier werden rhetorische Kompetenzen an zahlreichen Stellen erwähnt bzw. gefordert, z.B. unter „Pragmatische Kompetenzen“, S. 123 ff, „Kommunikative Aufgaben“ S. 153 ff, oder auch die Einladung, „Präsentationen/Referate/Vorträge“ selbst zu produzieren S. 144).

⁸ In Italien ist das in der Ausbildung für Lehrpersonen bislang gar nicht vorgesehen.

⁹ Zur Situation in Polen siehe Grzeszczakowska-Pawlikowska (2016).

rhetorische Kompetenzen im Kontext der Selbst- und Fremdwahrnehmung von DaF-Lernenden zu beleuchten.

2.1. Akademischer Diskurs und Verständlichkeit

DaF-Studierende an Hochschulen und Universitäten stehen vor der Herausforderung, nicht nur sprachliche Barrieren zu überwinden, sondern auch die Fähigkeit zu entwickeln, sich im akademischen Diskurs angemessen auszudrücken. Dies erfordert eine umfassende Beherrschung der Sprache in all ihren Facetten, einschließlich des Lexikons, der Syntax und anderer linguistischer Parameter. Darüber hinaus ist es entscheidend, komplexe Gedanken strukturiert und überzeugend zu formulieren und sie verständlich sowie zielgruppenorientiert zu präsentieren. Das umfasst sowohl schriftliche als auch mündliche rhetorische Fähigkeiten, betrifft auf der Seite der Rezipierenden schriftliche Kompetenzen im Leseverstehen und mündliche im Hörverstehen.

An dieser Stelle soll auf die Untersuchungen in den 1970er und 1980er Jahren über Kriterien von Verständlichkeit schriftlicher Texte durch die Forschungen von Langer, Schulz von Thun und Tausch (1975) sowie Groeben (1976) verwiesen werden. Die in diesem Zusammenhang definierten vier Kriterien (Struktur und Ordnung, Einfachheit des Ausdrucks, Auswahl der Information, Anregung und Interesse) wurden in einem zweiten Schritt (Allhoff 1984) um weitere drei für den mündlichen Vortrag ergänzt: visuelle Hilfsmittel, hörergerechte Darbietung und Sprechweise. Diese Merkmale sind entscheidend für das Verstehen und die Interpretation der gesprochenen Äußerungen und tragen zur Optimierung der kommunikativen Effektivität bei (siehe auch Allhoff, Allhoff 2014, S. 62-73).

Entsprechend liegt in der im Folgenden besprochenen Studie der Fokus auf der mündlichen Äußerungskompetenz der Studierenden, wobei vor allem auf hörerorientierte Darbietung und Sprechweise geachtet wird. Dabei werden auch Aspekte wie Sprechgeschwindigkeit, Pausensetzung, Artikulationspräzision, Sprechvariation und die sinnstützende Verwendung von Intonationsmustern berücksichtigt.¹⁰ Als weiterer wichtiger Faktor wird die Art des Vortrags (frei gesprochen, gelesen oder auswendig vorgetragen) betrachtet, die sich wiederum in der Sprechweise widerspiegelt. Im Zusammenhang mit der Untersuchung wird daher den paraverbalen Merkmalen, v.a. der intonatorischen Fokusmarkierung als Verständlichkeitskriterium eine entscheidende Rolle beigemessen.¹¹

¹⁰ Dieser Teil der Untersuchung befindet sich noch in der Auswertung.

¹¹ Siehe dazu auch Büring 2006 und Kapitel 4.3. dieses Beitrags.

2.2. Sprechwirkung nicht-deutscher Vortragender

Die Sprechwirkungsforschung fungiert als theoretischer Rahmen zur Analyse, wie Sprache auf die Rezipierenden wirkt und welche Mechanismen dabei eine Rolle spielen. Stock (1991, S. 23) charakterisiert das Forschungsfeld der Sprechwirkungsforschung als eine komplexe Untersuchung des Perzeptions-, Rezeptions- und Reaktionsprozesses, der sowohl die Aufnahme als auch die Bewertung aller kommunikativ relevanten Elemente umfasst, wobei neben der sprachlichen Kommunikation auch visuelle Reize einbezogen werden. Die vorliegende Studie fokussiert primär die phonetische Wirkung, insbesondere die Auswirkungen von sprechstimmlichen Parametern, und sekundär die visuellen Elemente (Körpersprache).¹²

Für die Rezeption des gesprochenen Wortes durch die Zuhörenden sind aufeinanderfolgende Phasen erforderlich: zunächst die Analyse des akustischen Signals und die Identifikation der kommunikativ relevanten Merkmale. Daraufhin erfolgt die Zuschreibung ihrer Bedeutung und Funktion. Schließlich interpretieren die Zuhörenden die Äußerung unter Berücksichtigung des situativen Kontextes und erschließen deren kommunikativen Sinn. Diese kognitiven Verarbeitungsprozesse bilden die Grundlage für die Reaktion auf die Äußerung. Um zu verstehen, wie unser Hörsystem und Gehirn komplexe auditive Informationen verarbeiten, ist es wichtig, ihre bemerkenswerten Fähigkeiten zu erkennen:

Perceiving, processing and extracting of relevant information from the complex, multi-layered and variable ensemble of acoustic data are remarkable features of our auditory system and brain. We perceive and understand the stream of speech utterance (in form of acoustic waves) despite the variability and complexity issues at both linguistic and non-linguistic levels. Furthermore, we are able to extract relevant information out of the mass of input data as well as to reconstruct the information in case of a noisy or incomplete data stream. (Lapteva 2011, S. 12)

Die Analyse der Sprechwirkung gewinnt besonders an Relevanz im Kontext von DaF-Lernenden, die nicht in ihrer Muttersprache kommunizieren.

2.2.1. Verständlichkeit vs. Sympathiewirkung

Fremdsprachige Interferenzen führen oft zu vorzeitigen Schlussfolgerungen über Kompetenzeindruck, Verständlichkeit und Sympathiewirkung von Sprechern, was wiederum soziale Beziehungen und Interaktionen im Alltag beeinflussen kann. Eine von Kaunzner (2015, 2018) durchgeführte

¹² Da die Referate sowohl in Präsenz als auch online gehalten worden waren, liegt der Fokus auf stimmlich-sprecherischen Kriterien; lediglich vier allgemeine Fragen betreffen die nonverbale Wirkung (siehe Kap. 4).

mehrphasige Fallstudie mit italienischsprachigen Studierenden der deutschen Sprache verdeutlicht die Rolle der Aussprache von DaF-Lernenden und insbesondere die Wirkung des italienischen Akzents. Gegenstand waren Audio-Aufnahmen vorgelesener deutschsprachiger Nachrichtentexte durch italophone Studierende, die von Hörergruppen aus unterschiedlichen Kulturkreisen (Italien, Deutschland und Polen) bewertet wurden. Die Ergebnisse lassen sich wie folgt zusammenfassen.

Die Wahrnehmung und Bewertung der fremdsprachlichen (italienischen) Interferenzen zeigt interessante Unterschiede zwischen den verschiedenen Gruppen. Betrachtet man zunächst die italienische, so nehmen diese Personen ihren (eigenen) Akzent besonders auf der Lautebene wahr. Prosodische Interferenzen wie Abweichungen in Betonung, Rhythmus oder Intonation fallen ihnen weniger auf, und werden interessanterweise sogar in Bezug auf die Verständlichkeit von der italienischen Gruppe eher positiv bewertet. Auch bei der polnischen Gruppe korreliert die Verständlichkeit stärker mit Lautfehlern, während die deutsche Gruppe eine abweichende Prosodie sensibler wahrnimmt und als deutliche Verständnisbarriere empfindet.

In Bezug auf die Sympathiewirkung zeigt sich ein gegensätzliches Bild. Fremdsprachliche Interferenzen in der Prosodie der italophonen DaF-Lernenden beeinträchtigen das Sympathieempfinden bei deutschen und polnischen Hörerinnen und Hörern kaum, selbst wenn die Verständlichkeit abnimmt. Bei der italienischen Hörergruppe ist dies nicht der Fall. Sie reagiert empfindlicher auf prosodische Abweichungen hinsichtlich der Sympathiebeurteilung.

Dieses Ergebnis erstaunt umso mehr, als die Sympathiewirkung von der Verständlichkeit unabhängig scheint. Es bietet sich folglich an, nach kulturspezifischen Interpretationstendenzen zu fragen und die Bedeutung des fremden Akzents in interkulturellen Kommunikationssituationen sowie den Einfluss von Stereotypisierungen und Vorurteilen näher zu untersuchen.

Die hier beschriebene Studie greift die rhetorische Sprechwirkung in einem neuen Kontext auf und untersucht diesbezügliche Veränderungen über einen Zeitraum von knapp zwei Jahren (T1 im Studienjahr 2019/20 und T2 im Studienjahr 2021/22), wobei Selbst- und Fremdwahrnehmung sowie die spezielle Situation der Online-Präsentation beleuchtet werden.

2.2.2. Selbst- und Fremdwahrnehmung

Die Erforschung der Selbst- und Fremdwahrnehmung von Vortragenden ist ein multidisziplinäres Gebiet, das in erster Linie Erkenntnisse aus Psychologie, Neuropsychologie und Linguistik verbindet. Außerdem ist dabei zwischen objektiven und subjektiven Situationsfaktoren zu unterscheiden (Bose, Hannken-Illjes 2017, S. 511).

Psychologische Studien beleuchten die kognitive Verarbeitung von Sprache und die sozialen und kulturellen Einflüsse auf die Wahrnehmung der Rezipierenden. Hier spielen Vorurteile, Stereotypen und individuelle Erfahrungen eine wichtige Rolle.

Neurophysiologische Betrachtungsweisen beziehen sich auf die Umwandlung der eingehenden Schallsignale in neuronale Signale im Gehirn. Dabei interagieren verschiedene Hirnregionen, um die akustischen Informationen als bedeutungsvolle Sprache zu decodieren und zu verarbeiten, also Phoneme zu erkennen, Wörter zu klassifizieren und Bedeutung zu extrahieren. Das betrifft neben der reinen Inhaltsseite auch affektive bzw. emotionale Informationen, die in erster Linie durch die Prosodie markiert werden; weiter sind sprecherspezifische Elemente wie Alter, Geschlecht, Sprache/Dialekt/Akzent und körperliche Kondition von Relevanz (Lapteva 2011, S. 13).

Linguistische Analysen hingegen konzentrieren sich auf stimmliche, sprecherische und sprachliche Merkmale, die die Wahrnehmung beeinflussen. Dazu gehören Aspekte wie Akzent, Artikulation, Sprechmelodie, Sprechtempo und Sprech- und Sprachstil. Es werden automatisch Rückschlüsse auf Identität sowie soziale Wahrnehmung der Kommunizierenden gezogen (Fremdwahrnehmung). Weiter ist die Art und der Grad der Beeinflussung dieser Merkmale darauf, wie sich die Sprechenden selbst in Bezug auf ihre Sprach- bzw. Sprechkompetenz rezipieren von Interesse (Selbstwahrnehmung); für die hier besprochene Studie ist dieser letzte Aspekt wichtig.

2.3. Vermittlung rhetorischer Kompetenzen an der Unimore

Wie geht die Universität Modena und Reggio Emilia derzeit mit den Herausforderungen in Bezug auf rhetorische Kompetenzen von DaF-Lernenden um? Im Kurs „Lingua, cultura e istituzioni dei paesi di lingua tedesca“ des Bachelor-Studiengangs LCE (Lingue e culture europee) und im Kurs „Intercultural Communication and Language Variation – German Language“ des Master-Studiengangs LACOM (*Languages for Communication in International enterprises and organizations*) hat seit Jahren jeweils ein Modul zur Angewandten Rhetorik seinen festen Platz. Ein Blick auf die Beschreibung des Studiengangs LACOM verdeutlicht dessen professionelle Einbettung:

Il corso di laurea magistrale in *Languages for communication in international enterprises and organizations* prepara professionisti flessibili i quali, grazie alle solide competenze in almeno due lingue straniere europee e alle buone conoscenze di base in ambito giuridico ed economico, sono in grado di gestire con consapevolezza critica la comunicazione orale e scritta con interlocutori di

diverse lingue e culture, nel contesto di aziende, amministrazioni pubbliche e organizzazioni internazionali.¹³

Die Aufgaben im Arbeitskontext von den zitierten „Fachkräften“ im Bereich Öffentlichkeitsarbeit umfassen verschiedene Tätigkeiten. Zu ihren Hauptaufgaben gehört es unter anderem, Präsentationen zu halten oder diese in schriftlicher Form über die Aktivitäten des Unternehmens oder der Verwaltung vorzulegen, die sowohl intern als auch extern genutzt werden. Weiter führen sie mündliche Vermittlungstätigkeiten (Dolmetschen) von und zu den studierten Fremdsprachen durch.

Selbst der Bachelor-Studiengang LCE zielt darauf ab, angemessene Kenntnisse und kommunikative Fähigkeiten in mindestens zwei Fremdsprachen zu vermitteln. Er soll die Studierenden mit kommunikativen, interlinguistischen und interkulturellen Kompetenzen für die Ausübung von Berufen der mündlichen und schriftlichen Sprachmittlung sowie das Management der Kommunikation in mehrsprachigen Umgebungen ausstatten.

Die folgenden Ausführungen beziehen sich auf die Kurse in deutscher Sprache: „Lingua, cultura ed istituzioni dei paesi di lingua tedesca“ (LCE, 2. Jahr) und „Intercultural communication and language variation – German language“ (LACOM, 1. Jahr). Im Rahmen des Unterrichts nimmt die Vermittlung rhetorischer Fähigkeiten neben sprachlichen Kompetenzen eine zentrale Rolle ein; das betrifft sowohl die Rede- als auch die Gesprächsrhetorik.

So liegt im zweiten Jahr des Bachelor-Studiengangs LCE der Fokus auf rhetorischen Grundlagen und Grundelementen der Rederhetorik, dem sach- und hörerbefugten Informieren und Überzeugen, der Vorbereitung in Teamarbeit und dem freien Halten einer videoaufgezeichneten Kurzrede in Form eines Gruppenreferats (5-10 Minuten pro Person) mit Hilfe von

¹³ Vgl. <https://www.unimore.it/it/didattica/corsi-di-studio/languages-communication-international-enterprises-and-organizations>, [11.08.2024]: Der Masterstudiengang *Languages for communication in international enterprises and organizations* bereitet flexible Fachkräfte vor, die dank solider Kompetenzen in mindestens zwei europäischen Fremdsprachen und guter Grundkenntnisse in den Bereichen Recht und Wirtschaft in der Lage sind, die mündliche und schriftliche Kommunikation mit Gesprächspartnern verschiedener Sprachen und Kulturen im Kontext von Unternehmen, öffentlichen Verwaltungen und internationalen Organisationen kritisch zu führen. (Übersetzung UK).

Auf der Homepage des Bachelor-Studiengangs LCE heißt es: “Gli strumenti linguistici e culturali che il corso fornisce consentono ai laureati di operare proficuamente nel settore della comunicazione multilingue, sia come traduttori che come redattori e consulenti linguistici, principalmente per imprese e organizzazioni internazionali, ma anche per case editrici e agenzie pubblicitarie e turistiche.” (Übersetzung UK: Das sprachliche und kulturelle Rüstzeug, das der Studiengang vermittelt, ermöglicht es den Absolventen, gewinnbringend im Bereich der mehrsprachigen Kommunikation zu arbeiten, sowohl als Übersetzer als auch als Sprachredakteure und Berater, vor allem für internationale Unternehmen und Organisationen, aber auch für Verlage, Werbe- und Tourismusagenturen.) (<https://www.unimore.it/it/didattica/corsi-di-studio/lingue-e-culture-europee>) [11.08.2024].

Stichwortkonzepten und der Beantwortung unvorhersehbarer Fragen aus dem Publikum (Fragerunde). In einer abschließenden Besprechung mit der Lehrkraft reflektieren die Studierenden über die persönliche Redeleistung während ihres Referats und der Fragerunde, wobei sie sich mit ihren eigenen Videoaufnahmen befassen. Diese Vorgehensweise ist aus didaktischer Sicht besonders wertvoll und entspricht den Taxonomiestufen von Bloom in der überarbeiteten Version von Anderson und Krathwohl (2001, S. 28-29), bei der vom Faktenerwerb ausgehend über die konzeptuelle und prozedurale Ebene die metakognitive Niveaustufe erreicht werden soll.

Diese vermittelten Grundlagen der Rederhetorik umfassen die Kenntnis und Anwendung verschiedener Redestile (informierend vs. überzeugend), Kriterien für Verständlichkeit, das Verständnis für die Bedeutung nonverbaler und paraverbaler Elemente etc. Mit Hilfe eines transparenten, kriteriengestützten Beobachtungs- bzw. Beurteilungsbogens wird die Vortragsleistung der Studierenden individuell bewertet (Apel, Méndez, von Laguna 2015, S. 109-112). Nicht zuletzt lernen die Studierenden die Grundsätze der Slide-Gestaltung und ihres Einsatzes während der Präsentation sowie das Schreiben eines Handouts. Wie Forster (1997, S. 296) festhält, gilt der Wissenserwerb zur Sprech- und Hörhandlung im Kontext des Referierens als unabdingbar, darf aber nicht als isolierte Unterrichtseinheit betrachtet werden.

Der Kurs des Master-Studiengangs LACOM vertieft die Inhalte des Bachelor-Studiums und legt anhand verschiedener Gesprächssituationen den Schwerpunkt auf internationale Rhetorik sowie die Kommunikation im Berufskontext (Bewerbungsgespräch, Verhandlung, Mitarbeitergespräch, Meeting-Management, Beratungs- und Verkaufsgespräch). Hierbei wird besonderer Wert auf aktives Zuhören und Argumentation gelegt. Auch diese Studierenden halten ein Gruppenreferat, was nach denselben Kriterien beurteilt wird und Teil der mündlichen Note ist.¹⁴

3. Das Korpus *Studentische Referate in der FS Deutsch*

Das Projekt der Datenbank ist Teil einer größeren Projektgruppe und soll zunächst vorgestellt werden (3.1.), bevor das der vorliegenden Longitudinalstudie zugrundeliegende Korpus beschrieben wird, d. h. die

¹⁴ Der Besprechung liegt ein den Studierenden bekannter Bewertungsbogen zugrunde: 1. Inhalt (Einstieg, Schluss, Gliederung, Verständlichkeit); 2. Sprache und Sprechausdruck (Aussprache, Intonation, Grammatik, Satzbau, Einsatz der Redemittel); 3. Vortrag (Publikumskontakt und -bezug, Nonverbales: Körpersprache, Gestik, Blickkontakt, Bewegung...; Vortragstechnik, Bezug zu den Slides, Slide-Wechsel); PowerPoint und Handout (Layout, sprachliche Gestaltung, Stil). Weiter fließen in die Beurteilung der Vortragsmodus (frei, auswendig, gelesen), der Stil (mündlich vs. schriftlich), das Zeitmanagement, die inhaltliche Relevanz und der Kursbezug ein.

Datensammlung und ihre Aufbereitung mit Hilfe des Tools OCTRA zur Transkription und Bearbeitung der Daten, die anschließend in die Datenbank BAS CLARIN Repository eingespeist wurden (3.3.).

3.1. Wissenschaftliche Einbettung des Korpus

Die Projektgruppe *Comunicazione Accademica e Professionale e Disseminazione della Conoscenza nell'era digitale* (Akademische und berufliche Kommunikation und Wissensverbreitung im digitalen Zeitalter, im Folgenden „CAP Deutsch“) ist am Dipartimento di Studi Linguistici e Culturali der Universität Modena und Reggio Emilia angesiedelt.¹⁵ Das Gesamtprojekt umfasst mehrere Studiengruppen, die diverse Forschungsbereiche und Sprachen (neben Englisch auch Deutsch, Französisch und Spanisch) abdecken. Es konzentriert sich auf die Analyse verschiedener schriftlicher Texte aus akademischen und beruflichen Kontexten, wobei das Projekt CAP Deutsch besonders die mündliche Sprache untersucht.

CAP Deutsch zielt darauf ab, die gesprochene Sprache italienischer Studierender anhand von Präsentationen (Referaten) in deutscher Sprache (L2) (vgl. 2.3.) zu digitalisieren, in die Datenbank BAS CLARIN Repository einzuspeisen und zu analysieren. Von den Referaten wird jeweils der Beginn des vorbereiteten Teils und der anschließenden Diskussion aufgezeichnet. Alle Probanden stimmen der Aufzeichnung und Verwendung des Materials für akademische Zwecke zu. 2024 zählt die Datenbank 450 digitalisierte und transkribierte Präsentationen und Fragerunden (vgl. 3.3.).

Für die vorliegende Zweipunkt-Longitudinalstudie, die in diesem Beitrag behandelt wird, wurde folgendes Kriterium für die Auswahl der Daten angewandt: Es handelt sich um Präsentationen derselben Studierenden im Abstand von knapp zwei Jahren (T1 und T2). Dieses Profil hatten 14 Studierende, so dass sie als Probanden ausgewählt wurden.

Zum Arbeitsteam zählten zudem zwei Praktikantinnen und eine Mitarbeiterin des Sprachenzentrums.¹⁶ Die Vorbereitung der Studie gliederte sich in die Phasen Datenerhebung, Aufbereitung des Materials und Transkription.

3.2. Datenerhebung, Aufbereitung des Materials und Transkription

Die Videoaufnahmen der unter 2.3. erwähnten Referate und ihrer anschließenden Diskussion dienen der Erstellung der im Projekt verwendeten

¹⁵ Im Jahr 2022 setzte sich die Forschungsgruppe neu zusammen.

¹⁶ Für alle technischen Fragen bezüglich der Sprachkorpora und des BAS CLARIN Repository wurde Christoph Draxler, wissenschaftlicher Mitarbeiter des Instituts für Phonetik und Sprachverarbeitung (IPS) an der Ludwig-Maximilian-Universität München, hinzugezogen.

Datenbank als Teil des BAS CLARIN Repository¹⁷. Übergeordneter Zweck der Videoaufnahmen ist es, diese für linguistische Forschungszwecke zugänglich zu machen. Um auch zu einem späteren Zeitpunkt schnell und einfach auf die wichtigsten Informationen über die Sprecher schließen zu können, enthalten die Dateinamen anonymisierte Informationen über das Studienjahr, den Studiengang, das Geschlecht, die in der Präsentation genutzte(n) Sprache(n), das Thema, die Vortragsmodalität, den Sprechstil sowie die Sprechersequenz in der jeweiligen Referatsgruppe.

Vor diesem Hintergrund entstehen mithilfe der Software Avidemux¹⁸ kurze Videoausschnitte der studentischen Referate und der anschließenden freien Diskussion (jeweils eine Minute des Redebeginns). Die Software Audacity¹⁹ ermöglicht es anschließend, die Audiospur aus dem Video als WAV-Datei zu exportieren. Diese beiden Arbeitsschritte gewährleisten nicht nur eine Verbesserung der Audioqualität²⁰ bei geringerer Dateigröße, sondern verleihen dem Material auch die notwendigen Eigenschaften zur Aufnahme in die Datenbank. Dieses Format ermöglicht dazu die orthographische Transkription mit dem Tool OCTRA, auf das in Folge näher eingegangen werden soll.

OCTRA²¹ ist ein webbasierter Transkriptionseditor, der ohne eigene Softwareinstallation in den meisten modernen Browsern verwendet werden kann. Er bietet die Möglichkeit, auf in einer Datenbank gespeicherte Audioaufnahmen zuzugreifen, sodass zeitlich parallel und geographisch verteilt transkribiert werden kann. Die Datenbank bietet den Transkribierenden immer die nächste verfügbare Audiodatei an, ohne dass Daten manuell kopiert oder geöffnet werden müssen. Die Speicherung der Transkripte erfolgt automatisch nach Beendigung eines Transkriptionsvorgangs. Diese Vorgehensweise erleichtert die Zusammenarbeit unter den Projekt-Mitarbeitenden, die sich jeweils mit ihren persönlichen Anmeldedaten in OCTRA einloggen können, was für deutliche Zeitersparnis sowie Flexibilität sorgt.²² Im Falle mehrerer Mitarbeitenden, die sich sowohl der

¹⁷ Die Kriterien zur Aufnahme in die Datenbank, die unter anderem den in Folge dargestellten Arbeitsschritten zugrunde liegen, sind unter www.bas.uni-muenchen.de verfügbar.

¹⁸ <https://avidemux.sourceforge.net/> [14.05.2024]

¹⁹ <https://www.audacityteam.org/> [14.05.2024]

²⁰ Bis Juni 2021 erfolgte die Aufnahme mit dem iPhone bzw. im Fall von Online-Referaten anhand von Bildschirmaufnahmen. Wegen häufig mangelhafter Audioqualität wird im Zuge der laufenden Aufnahmen auf das Wireless-Mikrofon Saramonic Blink500 B6 mit USB-C Eingang zurückgegriffen.

²¹ Die folgenden Ausführungen zur Funktionsweise und den Eigenschaften von OCTRA basieren auf Pömp, Draxler (2019) sowie auf dem Online-Tutorial „OCTRA – Transcribing Speech in your Browser“ von Christoph Draxler am 8. Oktober 2021. Wie Pömp, Draxler (2019) schreiben, ist OCTRA ein Web-basierter Editor zur effizienten Anfertigung zeitalignierter orthographischer Transkripte von Sprachaufnahmen.

²² Alternativ dazu ist es auch möglich, im lokalen Modus einzelne Audiospuren direkt im Browser zur Transkription hochzuladen.

Rohtranskription als auch dem zweiten Arbeitsschritt der Qualitätskontrolle widmen, ist keine weitere Absprache bzw. Aufteilung der zu transkribierenden Dateien notwendig.

Zusätzlich bietet OCTRA zahlreiche Funktionen zu einer unkomplizierten Handhabung und Anfertigung der Transkription. In erster Linie soll hier die Möglichkeit zur Visualisierung der Audiodatei hervorgehoben werden, indem man die Datei in Segmente unterteilt. Die hierfür eingesetzten Trennungszeichen (siehe senkrechte Markierungen in Abbildung 1) fallen üblicherweise mit der Markierung von Pausen zusammen. Der Detaileditor ermöglicht deren gezielte Bearbeitung.

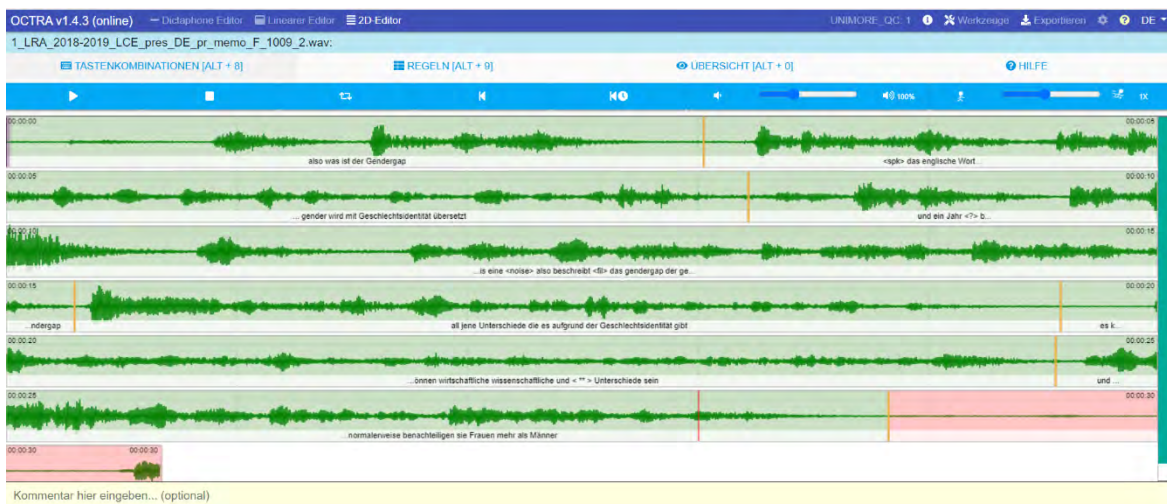


Abbildung 1
Transkription in OCTRA – Gesamtansicht.

OCTRA verfügt über eine Auswahl an Markersymbolen zur Verdeutlichung paraverbaler Phänomene im Transkript.²³ Die transkribierten Wörter bzw. zusätzlichen Anmerkungen anhand der Markersymbole werden automatisch zentral im jeweiligen Segment angeordnet (Abbildung 2, Detaileditor).

²³ Vor der Transkription wird eine umfassende Legende erstellt, in der die Symbole mit bestimmten Bedeutungen verknüpft sind. Diese Vorbereitung empfiehlt sich, da die von OCTRA bereitgestellten Icons grundsätzlich beliebig eingesetzt werden können. Es gibt hier keine Konventionen wie beispielsweise GAT-2, und es existiert kein Handbuch, das den Symbolen eine klare Bedeutung zuordnen würde. Die direkte Nutzung beschleunigt nicht nur den Arbeitsvorgang, sondern stellt auch die Uniformität der Transkriptionen sicher, was von den Mitarbeitenden im Projekt bestätigt wurde.



Abbildung 2
Transkription in OCTRA – Arbeit mit dem Detaileditor.

Die Integration des Tools WebMAUS ermöglicht bei automatischer Spracherkennung die korrekte Positionierung kleinerer linguistischer Einheiten wie Morpheme, Silben oder Phoneme, was zu tiefgreifenderen Folgeanalysen einlädt.²⁴ Eine manuelle Bearbeitung in diesem Zusammenhang wäre nicht nur zeitlich aufwändiger und kostenintensiver, sondern wäre auch fehleranfälliger, wie Kisler, Schiel, Sloetjes (2012, S. 30) hervorheben. Als Voraussetzung für ein zufriedenstellendes Ergebnis unterstreichen die Autoren Audioqualität und Standardsprache.²⁵

Nach Fertigstellung der manuellen Rohtranskription mit OCTRA sowie gegebenenfalls der automatischen Positionierung der kleineren linguistischen Einheiten mit WebMAUS erfolgt die Sicherung in der OCTRA-Datenbank. Neben der Möglichkeit zur Bewertung der Audioqualität ist ein Kommentarfeld vorhanden, um die Zusammenarbeit in Hinblick auf die anschließende Qualitätskontrolle zu erleichtern. OCTRA gibt für diesen

²⁴ Wichtig ist hier anzumerken, dass OCTRA zwar bei Bedarf mit WebMAUS kombiniert werden kann, es sich aber um zwei separate Tools handelt. WebMAUS ist auf BAS in verschiedenen Versionen (Basic, General, mit Pipeline with(out) ASR) verfügbar und kann unter folgendem Link aufgerufen werden: <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic> [24.06.2024].

²⁵ Draxler (2022, S. 34-35) befasst sich explizit mit der Fehlerrate der automatischen Spracherkennung, die insbesondere durch die fünf Faktoren (Anzahl der Sprecher, Audioqualität, Sprachstil, Dialekt bzw. Akzent und Diskursdomäne) beeinflusst wird. Während der Transkription der studentischen Referate in der Fremdsprache Deutsch konnten diese Limitationen bestätigt werden.

Folgeschritt die Audiospur mit der bereits angefertigten Rohtranskription automatisch aus.

Die fertigen Transkripte werden auch nach der Qualitätskontrolle automatisch in der OCTRA-Datenbank gespeichert, können dort abgerufen und anschließend als orthografische Transkripte mit oder ohne Zeitstempel und den Protokolldaten des Transkriptionsvorgangs exportiert werden. Zur Ermöglichung eines breiten Spektrums an Bearbeitungs- und Analyseoptionen²⁶ ist die Ausgabe in verschiedenen Dateiformaten möglich, darunter AnnotJSON, TextGrid, oder das Tabellenformat CVS sowie das Mehrebenen-Annotationsformat BAS für symbolische und zeitbezogene Segmentdaten.

4. Pilotstudie zur rhetorischen Sprechwirkung studentischer Referate

In der Untersuchung wurden sowohl Eigen- als auch Fremdbewertungen vorgenommen, wobei die Fremdbewertung von unabhängigen Beurteilenden durchgeführt wurde, die die rhetorische Sprechwirkung der Studierenden analysierten. Die Pilotstudie nutzt das Subkorpus der Datenbank „Studentische Referate in der FS Deutsch“ und untersucht exemplarisch, wie sich die rhetorische Sprechwirkung über einen Zeitraum von knapp zwei Jahren verändert. Dabei werden sowohl Unterschiede in der Selbst- und in der Fremdwahrnehmung als auch zwischen Präsenz- und Online-Präsentationen in Bezug auf den Publikumskontakt analysiert.

4.1. Testdesign und Forschungsfragen

Die Longitudinalstudie befasst sich mit Videoaufnahmen von 14 Studierenden zu zwei Zeitpunkten (T1 und T2). Es handelt sich um Referate der Probanden in der Fremdsprache Deutsch während des zweiten Jahres ihres Bachelor-Studiums LCE (T1) und in einer ähnlichen Lehrsituation während des ersten Jahres ihres Master-Studiums LACOM (T2).²⁷ Im ersten Fall wird das Kompetenzniveau A2/B1 vorausgesetzt, im zweiten Fall B2/C1.²⁸ Gegenstand

²⁶ Draxler (2022) betont als Zweck der Transkriptionen die Folgeforschung und unterstreicht so die Bedeutung, die gewonnenen Daten gezielt durchsuchen, verarbeiten, analysieren sowie visualisieren zu können.

²⁷ Siehe Abschnitt 2.3. Nicht alle Studierenden von LCE setzen ihr Studium mit dem Studiengang LACOM fort. In diesem Jahrgang waren es 14 Personen, ihre Referate stellen das Subkorpus dar.

²⁸ Die Kompetenzstufen hängen von der individuellen Sprachenkombination der Studierenden ab. Deutsch wird im Bachelor-Studiengang auch als dritte Fremdsprache und ohne Vorkenntnisse angeboten.

sind die Videoaufzeichnungen der vorbereiteten Referate, hiervon jeweils die ersten 30 Sekunden²⁹.

Die rhetorische Kompetenzentwicklung wurde nach der Digitalisierung und Aufbereitung des Testmaterials anhand eines Online-Fragebogens eruiert und umfasst zwei Ansätze: eine diachrone Analyse, die den Lernfortschritt beurteilt, und eine synchrone Analyse, bei der Selbst- und Fremdwahrnehmung gegenübergestellt werden.

Da die Vortragssituation im ersten Fall (T1) in Präsenz stattfand und im zweiten Fall (T2) online³⁰, ist es möglich, auch dahingehend Veränderungen zu überprüfen. Die Fragen zur Analyse des Lernfortschritts konzentrieren sich auf vergleichbare Elemente der Stimme und Sprechweise und vier grundlegende Aspekte zur Körperbewegung, die in beiden Redesituationen erkennbar sind (siehe 4.2.). Der vorliegende Beitrag greift aus der Fülle der Daten³¹ drei Fragen auf, die als folgende Hypothesen formuliert werden können:

- H1: In Bezug auf die Sprechwirkung kann man von einer Konsolidierung und weiteren Verbesserung der rhetorischen Kompetenzen ausgehen (T1 im Vergleich zu T2).
- H2: Die subjektive Einschätzung (Selbstwahrnehmung) und die objektive Wahrnehmung durch externe Evaluierende (Fremdwahrnehmung) unterscheiden sich dahingehend, dass die Externen kritischer sind.
- H3: Studierende haben bei online durchgeführten Referaten mehr Schwierigkeiten als bei Referaten in Präsenz, Kontakt mit dem Publikum herzustellen und zu halten.

4.2. Methode: Online-Umfrage

In diesem Abschnitt werden die Evaluationsgruppen und der Fragebogen vorgestellt, auf dem die Online-Umfrage zur Datenerhebung von Juni 2021 basiert und aus dem später aussagekräftige Ergebnisse ausgewählt und diskutiert werden. 32 Personen wurden bei dieser Datenerhebung mittels Google Forms³² zu 28 Videoaufnahmen befragt (T1: 14 Videos; T2: 14 Videos). Die Videos, die zu diesem Zweck in einem Ordner auf Google Drive freigegeben wurden, konnten beliebig oft wiedergegeben und pausiert werden.

²⁹ 30 Sekunden ermöglichen zwar einen ersten Gesamteindruck zum Vortrag, aber keine Beobachtungen des Aufbaus und der Redestruktur. Aus diesem Grund wird in der hier vorgestellten Studie nicht auf solche Aspekte eingegangen.

³⁰ Das zweite Referat entstand im Studienjahr 2021/22 zur Zeit der COVID-19 Pandemie, in der die didaktischen Aktivitäten ausschließlich online erfolgen konnten und die Implementierung digitaler Technologien einen neuen Höhepunkt erreichte.

³¹ Detailliertere Analysen und spezifische Fragestellungen werden in separaten, weiterführenden Studien bearbeitet und veröffentlicht.

³² Die Ergebnisse können in Form von Excel-Dateien exportiert werden.

Zur Eruierung der Eigen- und Fremdwahrnehmung gibt es drei Gruppen von Bewertenden: (A) 14 Studierende, die selbst die Referate gehalten haben (Selbsteinschätzung) und (B) 19 externe Evaluierende (Fremdeinschätzung). Gruppe B teilt sich wiederum in zwei Untergruppen: 8 Lehrpersonen (B-1) und 11 Studierende von LACOM, die zu T1 noch nicht an der Unimore eingeschrieben waren (B-2).³³ Diese Aufteilung der Gruppe B ist insofern relevant, da es sich bei B-1 um eine Expertenmeinung (Lehrpersonen) handelt.

Angesichts der eben dargelegten Unterteilung der Evaluationsgruppen A und B-1/B-2 basiert die Online-Umfrage auf zwei Fragebögen mit jeweils 24 identischen Fragen zu den Videoausschnitten (auf Deutsch und Italienisch). Die Vielzahl der angesprochenen Themen ermöglicht es später, Aspekte punktuell herauszufiltern und zu beantworten. Die beiden Fragebögen unterscheiden sich ausschließlich in Bezug auf einen zusätzlichen einleitenden Teil, der bei (A) etwas ausführlicher ausfällt und die Vorbereitung auf Referate und deren Umsetzung betrifft.³⁴

Die Fragebögen beinhalten einen Abschnitt zur persönlichen Sprachbiographie, um die jeweiligen Hintergründe der Teilnehmenden bezüglich der Variablen abzustecken, die den Prozess des Fremdspracherwerbs beeinflussen können. Darunter fallen Faktoren wie Alter, Geschlecht, Muttersprache(n) und Lernmotivation, aber auch Aspekte in Bezug auf die Lernumgebung sowie den Lernkontext (Diadori 2015, S. 2-12); im Fall der Lehrpersonen wird auch ihre Lehrerfahrung ermittelt.³⁵

Zum Profil der Evaluationsgruppen: Die Bewertenden der Gruppe A (14 Probanden, L1 Italienisch, weiblich n=13, männlich n=1) hatten die Sprachen Englisch und Deutsch als Studienfächer und waren zum Zeitpunkt der Umfrage zwischen 22 und 24 Jahre alt. 50 % der Probanden lernten seit ihrem Bachelorstudium an der Unimore Deutsch, was bedeutet, dass die erste Videoaufnahme (T1) in ihrem zweiten Lernjahr, die zweite Aufnahme (T2) in ihrem vierten Lernjahr stattfand. Die zweite Hälfte der Probanden gab an, vor dem Bachelorstudium an der Unimore bereits fünf Jahre an einem Gymnasium (*liceo*) in Italien Deutsch gelernt zu haben. 79 % verbrachten zwischen der ersten und zweiten Videoaufnahme einen Praktikumsaufenthalt oder ein Erasmus-Semester im deutschsprachigen Ausland (maximal 6 Monate). Außerhalb des Studienkontextes wurde der regelmäßige Kontakt zu

³³ Der gegebene Unterschied der Teilnehmerzahl der Gruppen A und B legt nahe, die Ergebnisse ausschließlich in Prozent darzustellen (siehe Abschnitt 4.3.).

³⁴ In diesem Abschnitt des Fragebogens wurden die Probanden zu ihren Gewohnheiten, Strategien und persönlichen Einstellungen zum Thema Referate befragt, um sich als Folgefragestellung über die rhetorischen Kompetenzen hinaus mit didaktischen Verfahren und deren Verbesserung befassen zu können, mit Fokus insbesondere auf die drei zum Spracherwerb unbedingt erforderlichen Grundgrößen Sprachverarbeiter, Input und Antrieb (Klein, Dimroth 2003, S. 18).

³⁵ Die Probanden konnten im Fragebogen die Teile überspringen, die für ihre jeweilige Kategorie nicht relevant waren.

Muttersprachlern bei 15% mit „oft“, bei 35% mit „gelegentlich“ und bei 50% mit „fast nie“ angekreuzt, was jedoch laut eigener Angaben der Studierenden die individuelle Beschäftigung mit der deutschen Sprache durch Lesen, Filme etc. nicht ausschließt.

Die externen Evaluierenden (Gruppen B-1 und B-2) zeigen folgendes Profil: B-1 verfügt über DaF-Lehrkompetenzen im universitären und/oder schulischen Kontext. Es sind Personen mit deutscher (n=5) oder italienischer (n=2) bzw. deutscher sowie italienischer (n=1) Muttersprache. B-2 sind LACOM-Studierende mit L1 Italienisch, die die Fremdsprachen Deutsch und Englisch im Studium gewählt haben.

Die Fragen zu den Videos befassen sich mit der jeweiligen rhetorischen Sprechwirkung der Referate und mit daran geknüpften Elementen zum Sprechstil. Eine maßgebliche Rolle spielen dabei Interferenzen zwischen der Vortragssprache Deutsch und der Ausgangssprache Italienisch.³⁶ Der die Sprechwirkung betreffende Teil des Fragebogens gliedert sich in folgende Abschnitte:

- Gesamteindruck (Abbildung 3): Hier werden auf einer mehrstufigen Antwortskala (Likert-Skala) mit vier Items von „ja“ über „eher ja“ und „eher nein“ bis „nein“ die Verständlichkeit, der Sprechfluss, die Natürlichkeit, die Sicherheit, der Kompetenzeindruck, die Sympathiewirkung³⁷ und der Kontakt zum Publikum bewertet. Abbildung 3 zeigt diese Matrixfragen zum Gesamteindruck.

³⁶ Zur Wirkung des fremdsprachlichen Akzents siehe Kaunzner 2015, hier besonders S. 72.

³⁷ Für die vorliegenden Fragestellungen ist dieser Punkt sekundär; er wurde in Bezug auf eine weiterführende Analyse den fremden Akzent betreffend aufgenommen, auf den hier nicht eingegangen wird.

Gesamteindruck zur/zum Vortragenden / Impressione generale sul/la relatore/rice *

	ja / sì	eher ja / più sì che no	eher nein / più no che sì	nein / no
Verständlichkeit / comprensibilità	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Flüssigkeit / scorrevolezza	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Natürlichkeit / naturalezza	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sicherheit / sicurezza	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kompetenz / competenza	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sympathie / simpatia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kontakt zum Publikum / contatto con il pubblico	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 3
Mehrstufige Antwortskala zum Gesamteindruck.

- Vortragsart: Nach Angabe der beobachteten Vortragsart in Form einer Multiple-Choice-Frage (frei, gelesen, oder auswendig vorgetragen) mit anschließender Likert-Bewertung werden das Vorhandensein von Füllwörtern und der Stil des Vortrags (mündlich oder schriftlich)³⁸ bewertet.
- Stimme und Sprechweise: Eine fünfstufige Bewertungsskala dient der Analyse unterschiedlicher Kategorien (Stimmsenkungen, Variation und Betonung, Geschwindigkeit, Pausen, Artikulationspräzision, Sprechstimmlage) (Abbildung 4).

³⁸ Genauer Wortlaut: Ist der Stil mündlich? (nicht von schriftlicher Vorlage kopiert).

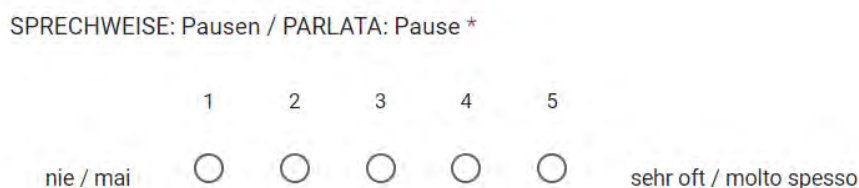


Abbildung 4
Bewertungsskala zur Sprechweise (Beispiel: Pausen).

- Sprachkenntnisse: Anhand einer Likert-Skala mit vier Items von „ja“ bis „nein“ werden Satz- und Wortbetonungen, Grammatik und Ausdruck und der fremdsprachige Akzent eruiert.³⁹
- Nonverbale Wirkung: Auch für diese Kategorie wird eine Likert-Skala mit vier Items zur Bewertung von Körperbewegungen, Gestikulation, Blickkontakt und Mimik herangezogen.

4.3. Ausgewählte Ergebnisse

Die nachfolgende Ergebnisdiskussion greift die drei in Abschnitt 4.1. formulierten Hypothesen auf: Man kann von einer Konsolidierung und/oder weiteren Verbesserung einmal erworbener rhetorischer Kompetenzen ausgehen (diachroner Ansatz; Hypothese 1). Im Vergleich zwischen Selbst- und Fremdeinschätzung zeigt sich letztere kritischer (synchroner Ansatz; Hypothese 2). Bei Online-Referaten ist es schwieriger als bei Referaten in Präsenz, Kontakt mit dem Publikum herzustellen und zu halten (Hypothese 3).

Hypothese 1 kann als verifiziert gelten, da keine der Evaluationsgruppen eine Stagnation oder Verschlechterung empfindet. Dies wird in Abbildung 5 deutlich: Die Bewertungen liegen im positiven Bereich, lediglich bei der Frage nach dem Publikumskontakt sind die Probanden selbstkritisch (diese Thematik wird weiter unten im Zusammenhang mit H3 aufgegriffen). Die Studierenden entwickeln sich im Laufe des Studiums (hier zwischen T1 und T2) in Bezug auf die rhetorischen Kompetenzen sogar weiter und gewinnen an Sicherheit. Neben den natürlichen Fortschritten, die den Studierenden mit zunehmender Studiendauer wichtige Möglichkeiten im Sprachgebrauch eröffnen, lässt sich die positive Gesamtwirkung ihres Vortrags auf gezielte Rhetorikschulung zurückführen, deren Erfolg langfristig erhalten bleibt.

Selbstverständlich hat der Erwerb von fachspezifischem rhetorischem Wissen an der Universität großes Gewicht; kommt es jedoch nicht zu einer didaktisch effizienten Umsetzung, d. h. praktischen Anwendung, werden

³⁹ Genauer Wortlaut: Klingt die Aussprache „Deutsch“?

langfristig gesehen kaum anhaltende Erfolge erzielt.⁴⁰ Eine der einleitenden Fragen für die Gruppe A lautete „Waren die Informationen und Übungen im Kurs zum Halten von Referaten auf Deutsch wichtig für Sie?“ und wurde zu 100 % mit „ja“ beantwortet.

Weiter kann eine intrinsische Motivation im Einklang mit positiver Verstärkung von außen einen wichtigen Beitrag zu erwünschten Langzeitresultaten leisten. Dies war u.a. durch die Vorbereitung des Kurzreferats in Teamarbeit gegeben, was für ein motivierendes Umfeld sorgte. Die Verifizierung der ersten Hypothese spricht daher dafür, dass die Studierenden in der Lage waren, das im Unterricht Gelernte anzuwenden und auch längerfristig erfolgreich umzusetzen.

Die zweite Fragestellung bezieht sich auf eventuelle Unterschiede zwischen Selbst- und Fremdwahrnehmung, wobei die Hypothese 2 davon ausgeht, dass die Externen kritischer beurteilen.

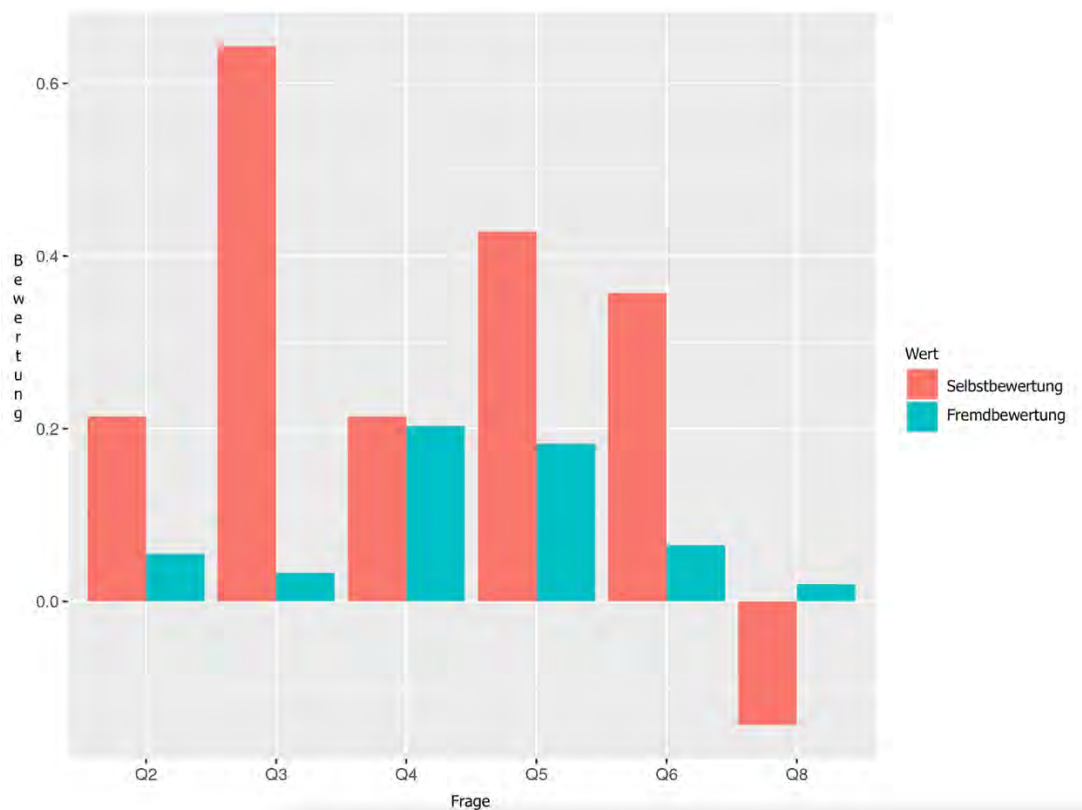


Abbildung 5

Durchschnittliche Abweichung T1 und T2:

Selbstbewertung (A) und Fremdbewertung (B-1 und B-2)

(Q2 = Verständlichkeit, Q3 = Sprechfluss, Q4 = Natürlichkeit, Q5 = Sicherheit, Q6 = Kompetenz, Q8 = Publikumskontakt).

⁴⁰ Wie sich die Studierenden mit den rhetorischen Grundlagen und Grundelementen der Rederhetorik praktisch auseinandersetzen, wird in Kapitel 2.3. erläutert. Siehe dazu auch Anderson und Krathwohl (2001, S. 28-29).

Wie aus den durchschnittlichen Abweichungen in Abbildung 5⁴¹ ersichtlich wird, sehen alle Gruppen Verbesserungen im Vergleich zu T1, jedoch nicht in gleichem Maße. Am besten urteilt die Gruppe A (Selbstbewertung), die Gruppen B-1 und B-2 (Fremdbewertung) sind kritischer. Lediglich bei Q8 (Kontakt mit dem Publikum)⁴² ist das nicht der Fall, hier glaubt die Probandengruppe sogar eine Verschlechterung wahrzunehmen. Besonders auffällig ist die positive Selbsteinschätzung der Vortragenden bei den Fragen zum Sprechfluss (Q3) und zum Kompetenzeindruck (Q6). Bei den Kriterien Verständlichkeit (Q2) und Sicherheit (Q6) ist ebenfalls eine Tendenz zur besseren Selbstbewertung zu erkennen, jedoch sind die Unterschiede zur Fremdbewertung hier weniger ausgeprägt.

Eine mögliche Erklärung für Unterschiede im Grad der positiven Bewertungen zwischen Selbst- und Fremdwahrnehmung kann zum einen sein, dass der weitere Kontakt mit der Sprache, sowohl im herkömmlichen Studienkontext als auch durch Erfahrungen wie Auslandsaufenthalte in Form von Praktika und Erasmus+ dazu beitragen, das Selbstbild der Vortragenden deutlich zu stärken. Zum anderen bestätigt der situative Kontext während der Vorbereitung und des Haltens der Referate das Streben nach „bedeutungsvoller Performance“ (vgl. Selinker 1972, S. 210). Die Folge daraus ist ein Handeln, das gerne als „kompetent“ eingestuft wird (Wisniewski 2018, S. 590).

Die Ergebnisse liefern wertvolle Erkenntnisse für die Selbstreflexion und das Feedback-Management im Rahmen von Rhetorikschulungen. Sie unterstreichen die Notwendigkeit, Vortragende für mögliche Wahrnehmungsunterschiede zu sensibilisieren und gezielte Maßnahmen zur Angleichung von Selbst- und Fremdwahrnehmung zu ergreifen.

Die folgenden Ausführungen analysieren nicht nur relevante Unterschiede zwischen Selbst- und Fremdwahrnehmung, sondern differenzieren zwischen B-1 und B-2, um zu sehen, ob die Expertengruppe (B-1) oder die Mitstudierenden (B-2) kritischer urteilen.⁴³ Dabei werden T1 und T2 gegenübergestellt. Abbildung 6⁴⁴ greift exemplarisch die Beobachtungsdimensionen Sprechfluss (Q3), Sicherheit (Q5) und Kompetenzwirkung (Q6) heraus.

Am klarsten zeigt sich ein positives Selbstbild von Gruppe A in Bezug auf die Beurteilung des Sprechflusses (Q3). Diese Gruppe schätzt sich selbst

⁴¹ Die x-Achse zeigt verschiedene Bewertungsdimensionen im Vergleich zu T1, wobei 0.0 keine Veränderung zeigen würde, 0.6 beispielsweise eine große Verbesserung darstellt. Auf der y-Achse sind die unterschiedlichen Fragen (Q) markiert, die Balken stehen für die Bewertung der Probanden (Gruppe A, jeweils linker Balken) und die der externen Bewertergruppen (Gruppen B-1 und B-2, jeweils rechter Balken).

⁴² Diese Thematik wird in Folge noch näher behandelt.

⁴³ Abschnitt 4.2. erläutert u.a. die Gründe für diese differenzierte Beobachtung.

⁴⁴ Auf der x-Achse sind in diesem Fall die möglichen Antworten der Likert-Skala nach Beurteilung der entsprechenden Kriterien aufgetragen: ja (3) – eher ja (2) – eher nein (1) – nein (0). Ein hoher Balken bedeutet eine positive Beurteilung, ein niedriger Balken einen negativen Wert.

als deutlich verbessert ein, während sich die Lehrpersonen (B-1) leicht optimistisch zeigen. Die Gruppe der Studierenden (B-2) will sogar einen Negativtrend erkennen. Um das zu erklären, muss man sich nochmals vor Augen halten, dass die Referate bei T1 in Präsenz und bei T 2 im Online-Modus stattfanden. Die Beurteilung des Sprechflusses könnte nun darauf hindeuten, dass die Kommilitonen (B-2) die Präsenzsituation als natürlicher und weniger ablenkend empfanden, während die Lehrpersonen (B-1) aufgrund ihrer Erfahrung und Professionalität die Leistungen besser relativieren können. Im Online-Modus ist das Publikum nur teilweise sichtbar, was zu weniger Ablenkung führt und eine stärkere Konzentration auf das, was man sagt und wie man es sagt, ermöglichen kann. Zudem lenken nervöse nonverbale Elemente bei der Beurteilung im Online-Format nicht ab, was die Selbsteinschätzung der Gruppe A positiv beeinflusst haben könnte.

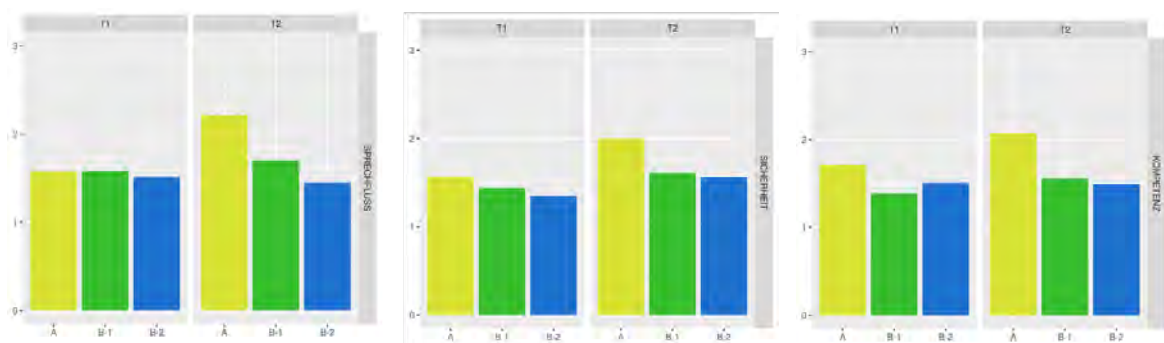


Abbildung 6

Bewertung zu T1 und T2 der unterschiedlichen Evaluationsgruppen in Bezug auf Sprechfluss (links), Sicherheit (Mitte) und Kompetenzwirkung (rechts).

Die in der Abbildung 6 erkennbaren Unterschiede in Bezug auf Kompetenz und Sicherheit lassen darauf schließen, dass die Studierenden ein erhöhtes Maß an Selbstvertrauen und eine optimistischere Einschätzung ihrer eigenen Fähigkeiten entwickelt haben. Auch die Fremdbewertung fällt diesbezüglich positiv aus, vor allem von Gruppe B-1, jedoch weit weniger stark als bei der Gruppe A. Diese Beobachtung sagt zum einen aus, dass die Fortschritte nicht nur subjektiv wahrgenommen, sondern auch objektiv so beurteilt werden. Die positive Veränderung in der Bewertung kann man somit als Indikator für den Kompetenzzuwachs und die gestärkte Sicherheit in der rhetorischen Praxis interpretieren, wobei die Fremdbewertenden zwar positiv, aber kritischer urteilen. Hypothese 2 ist hiermit auch verifiziert.

Die letzte hier zu beantwortende Frage rückt den Präsentationsmodus in den Fokus: T1 (Präsenz) und T2 (online); Hypothese 3 erwartet eine Verschlechterung in Bezug auf den Kontakt mit dem Publikum in der Online-Situation. Die studentischen Bewertenden (B-2) bemerken jedoch diesbezüglich keinen Unterschied, und für die Lehrpersonen (B-1) ist sogar

mehr Kontakt zum Publikum spürbar. Allein die Probanden (A) sind der Meinung, bei Online-Präsentationen weniger Kontakt zum Publikum herstellen und halten zu können als bei Präsentationen in Präsenz, obwohl sie, wie Abbildung 7 zeigt, diesen Punkt generell positiver bewerten als B-1 und B-2.

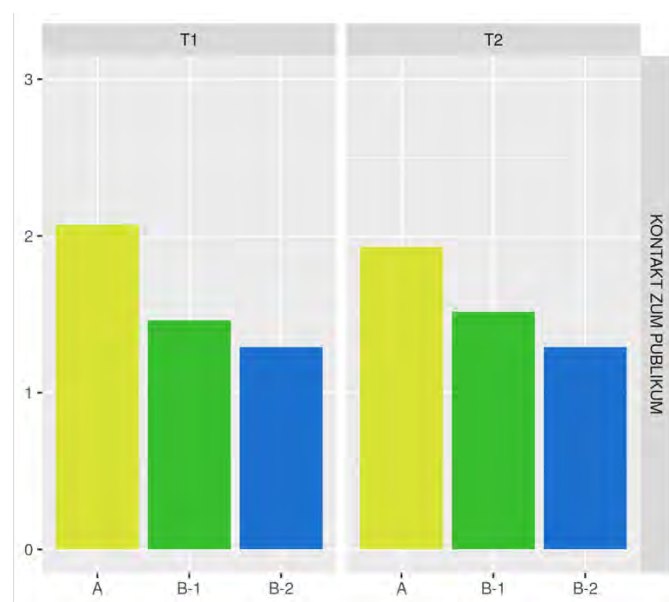


Abbildung 7

Bewertung zu T1 und T2 der unterschiedlichen Evaluationsgruppen in Bezug auf den Kontakt zum Publikum.

In diesem Zusammenhang ist ein Blick auf die Bewertungen nonverbaler Aspekte interessant: Gestik (Q22) und Mimik (Q23). Diese Elemente scheinen nicht notwendigerweise zu einem besseren Kontakt mit dem Publikum zu führen, obwohl diese physischen Ausdrucksmittel klarerweise in den Online-Präsentationen weniger ausgeprägt sind. Es wird bei T2 weniger Gestik wahrgenommen, die Mimik ist zwischen T1 und T2 (Abbildung 8) vergleichbar. Das deutet darauf hin, dass bei der Herstellung und Aufrechterhaltung des Publikumsbezugs in virtuellen Umgebungen andere Faktoren eine größere Rolle spielen können. Im Kontext des Online-Unterrichts war explizit an Strategien gearbeitet worden, die sich auf entsprechende Kommunikationssituationen beziehen, beispielsweise darauf, den virtuellen „Blickkontakt“ mit der Kamera herzustellen.⁴⁵ Hiermit in Zusammenhang steht folgende Frage, die der Gruppe A gestellt wurde:

⁴⁵ Andere online-spezifische Kenntnisse und Fähigkeiten betreffen die Art der Wahrnehmung der oder des Vortragenden in virtueller Umgebung, wenn kein direkter Kontakt mit dem Publikum gegeben ist oder wenn nur Ausschnitte sichtbar sind (i.d.R. der Kopf und Oberkörper) bzw. das Bild gänzlich fehlt. Eine mögliche Problematik, auf die Vortragende vorbereitet werden sollten, betrifft außerdem die Bewältigung von technischen Problemen, wie eine Störung der Aufnahme und Übertragung.

„Denken Sie, dass Sie durch die Online-Modalität in Bezug auf Referate benachteiligt waren?“ Da die Antworten von 72 % der Studierenden „Nein“ lautet, bedeutet das, dass sich ein Großteil der Studierenden auf die veränderte Vortragsmodalität vorbereitet fühlt. Diese Erkenntnisse widerlegen die dritte Hypothese.

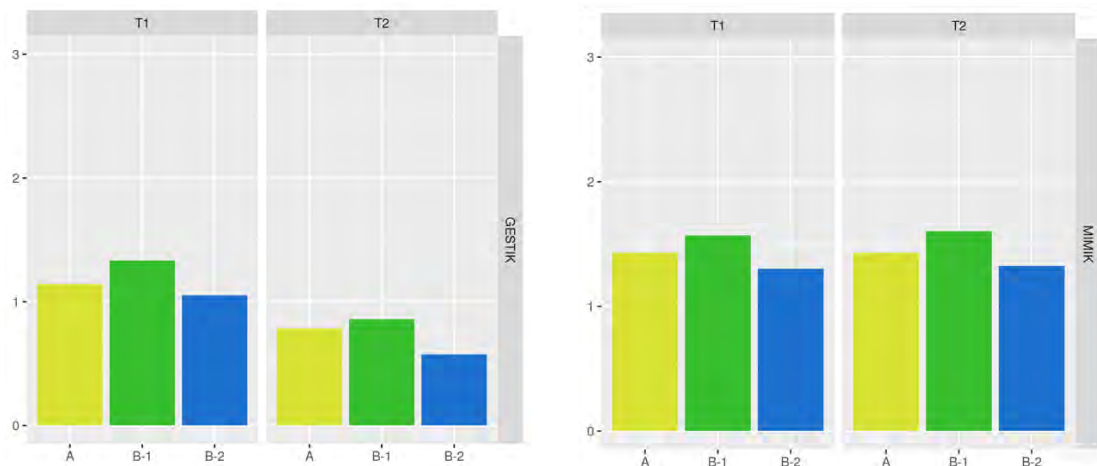


Abbildung 8

Bewertung zu T1 und T2 der unterschiedlichen Evaluationsgruppen in Bezug auf Gestik (links) und Mimik (rechts).

Abschließend soll gezeigt werden, wie die in diesem Abschnitt besprochenen Aspekte aufeinander wirken. Das Korrelationsdiagramm⁴⁶ (Abbildung 9) veranschaulicht die Beziehungen zwischen den Merkmalen Verständlichkeit (Q2), Sprechfluss (Q3), Natürlichkeit (Q4), Sicherheit (Q5), Kompetenzeindruck (Q6) und Kontakt (Q8) und vermittelt ein umfassenderes Verständnis davon, wie diese Faktoren zusammenwirken und die rhetorische Sprechwirkung eines Vortrags beeinflussen. Verbesserungen in einem Bereich können potenziell positive Auswirkungen auf andere Bereiche haben, was es ermöglicht, gezielte Interventionen und Trainingsmaßnahmen zu entwickeln, die nicht nur einzelne Aspekte verbessern, sondern auch den Gesamteindruck und die Kommunikationsfähigkeit der Vortragenden stärken. Darüber hinaus hilft es, potenzielle negative Wechselwirkungen zu identifizieren und Schulungsprogramme zu erstellen, um eine harmonische und konsistente Präsentation sicherzustellen.

⁴⁶ Die Farbintensität stellt in diesem Diagramm die Stärke der Korrelationen zwischen den Merkmalen dar: Hellere Farbtöne stehen für stärkere (positive) Korrelationen, während dunklere Farbtöne auf geringere (negative) Korrelationen hinweisen. Die visuelle Darstellung ermöglicht eine schnelle Identifikation signifikanter Zusammenhänge und Unterschiede. Diese Ergebnisse bieten wertvolle Einblicke in die wechselseitigen Einflüsse von Vortragseigenschaften und können als Grundlage für die Optimierung von Präsentationstechniken dienen.

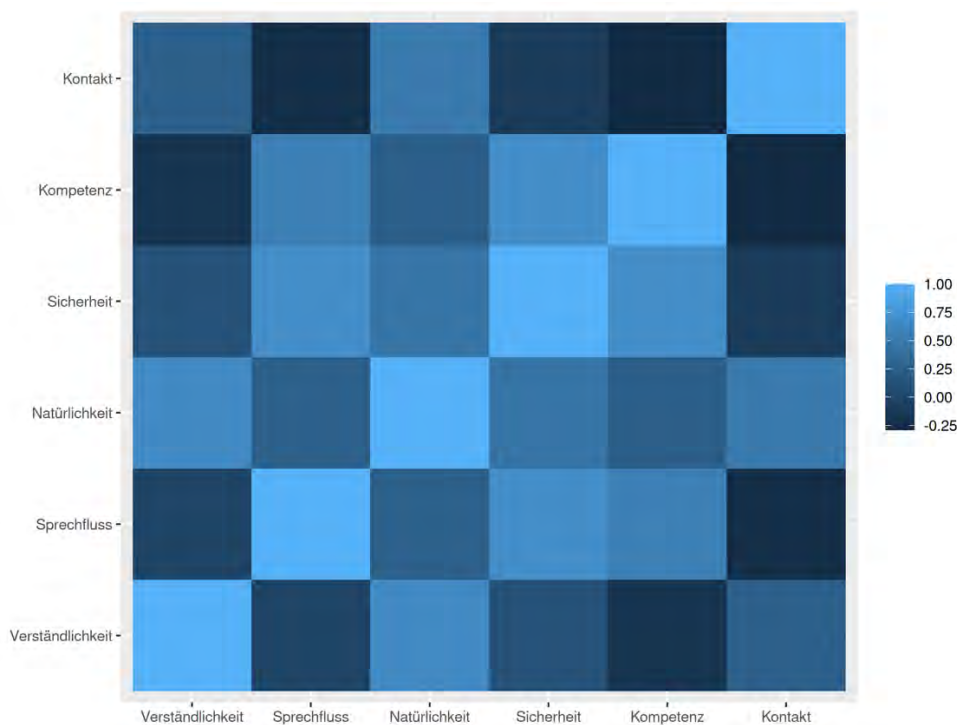


Abbildung 9

Analyse und Korrelation ausgewählter Aspekte aller Bewertungsgruppen
(A, B-1 und B-2 zusammengefasst).

Die stärksten positiven Korrelationen bestehen zwischen Verständlichkeit und Natürlichkeit, zwischen Sprechfluss und Sicherheit sowie zwischen Sicherheit und Kompetenzeindruck. Das deutet beispielsweise darauf hin, dass ein verständlicher Vortrag oft als natürlich und ein sicherer Vortrag als kompetent wahrgenommen wird. Zudem zeigt sich eine moderate positive Korrelation zwischen Kontakt und Verständlichkeit, was impliziert, dass ein guter Kontakt zum Publikum die Verständlichkeit des Vortrags fördert. Negative Korrelationen sind nicht signifikant vorhanden, was bedeutet, dass keine der untersuchten Merkmale stark gegensätzlich zueinander wirken.

5. Abschließende Überlegungen

Das Pilotprojekt hat gezeigt, wie effektives Arbeiten mit der Datenbank⁴⁷ „Studentische Referate in der FS Deutsch“ möglich ist, um die verschiedenen Aspekte der rhetorischen Sprechwirkung zu analysieren und nutzbare Erkenntnisse für die Sprachdidaktik zu gewinnen. Die Datenbank wurde und wird mit dem Ziel erstellt, generell für Forschungsarbeiten im Kontext der rhetorischen Sprechwirkung und anderen Bereichen der Angewandten Linguistik⁴⁸ zur Verfügung zu stehen.

Die hier herausgegriffenen Ergebnisse des Pilotprojekts offenbaren die vielschichtigen Auswirkungen rhetorischer Schulung und ihrer praktischen Anwendung auf die Entwicklung kommunikativer Kompetenzen im universitären Kontext.⁴⁹ Zunächst zeigt sich eine bemerkenswerte Verbesserung des Selbstbildes der Teilnehmenden, was auf ein gesteigertes Selbstvertrauen in ihre rhetorischen Fähigkeiten hindeutet. Dieses subjektive Empfinden wird durch objektiv messbare Fortschritte in der rhetorischen Kompetenz untermauert, was die Validität und Effektivität des Schulungsansatzes unterstreicht. Die Erkenntnisse legen daher nahe, dass die erworbenen Fähigkeiten einen substanziellen Beitrag zur erfolgreichen Teilnahme am akademischen Diskurs leisten. Die Transformation zu selbstbewussteren Vortragenden kann als Katalysator für eine aktivere und produktivere Beteiligung am akademischen Leben betrachtet werden.

Auch die Sprachdidaktik ist ein Feld, das dafür in Betracht gezogen werden sollte. Die Arbeit mit der Datenbank kann dazu beitragen, die Didaktik im Bereich DaF zu verbessern und als natürliche, wünschenswerte Folge die sprechsprachlichen Kompetenzen der Studierenden auf ein höheres Niveau zu heben. Würden beispielsweise häufige Fehlerkategorien im mündlichen Sprachgebrauch aufgezeigt werden, könnte man diese im Fremdsprachunterricht gezielt thematisieren.⁵⁰

Diese mehrdimensionalen Effekte unterstreichen die Bedeutung einer integrierten rhetorischen Ausbildung im Hochschulkontext. Sie verdeutlichen, dass die Vermittlung kommunikativer Kompetenzen weit über die bloße Verbesserung der Sprechfertigkeit hinausgeht und einen ganzheitlichen

⁴⁷ Vgl. Kapitel 3.

⁴⁸ Die Datenbank kann beispielsweise für folgende Gebiete von Interesse sein: Diskursanalyse, Gesprächsanalyse, Phonetik und Phonologie, Soziolinguistik, Korpuslinguistik.

⁴⁹ Dazu zählt auch der Sprachunterricht an der Universität, insbesondere durch Lektorate, die es den Studierenden erlauben, die bisher erworbenen Sprachkenntnisse praktisch anzuwenden.

⁵⁰ Ein induktiver Ansatz würde sich dabei besonders anbieten, der nicht nur zum eigenständigen Denken anregt, sondern, wie Michler (2015, S. 32) anmerkt, durch Selbsttätigkeit und die Arbeit mit konkreten, realen Sprachhandlungen auch die Motivation fördert. Motivation gilt bekanntlich als Schlüssel zum Erfolg – in diesem Fall, kombiniert mit fachspezifischem Wissen, zu einer erfolgreichen Kommunikation.

Beitrag zur akademischen, beruflichen und persönlichen Entwicklung der Studierenden leistet.

Zukünftige Forschung und curriculare Entwicklungen sollten daher die zentrale Rolle der rhetorischen Schulung in der Hochschulbildung weiter berücksichtigen und ausbauen. Anknüpfende Forschungsansätze könnten sich auf die folgenden Aspekte konzentrieren:

1. Identifikation besonders effektiver Komponenten der Rhetorikschulung: Mit einer Kontrollgruppe kann überprüft werden, in welchen Bereichen eine gezielte Schulung Mehrwert bietet.
2. Einsicht in die Lerntechniken und Präsentationsstile von Studierenden: Dabei könnten verschiedene Methoden zur Optimierung der Präsentationsergebnisse untersucht werden.
3. Einbeziehung anderer Sprachen und damit die Ausweitung der Datenbank.
4. Vertiefung interkultureller Fragestellungen, insbesondere:
 - a. Untersuchung der Sprechwirkung fremdsprachlicher Akzente: Wie beeinflussen fremdsprachliche Akzente die rhetorische Sprechwirkung?
 - b. Analyse kulturspezifischer Rezeptionsmuster: Welche kulturellen Unterschiede existieren in der Wahrnehmung rhetorischer Fähigkeiten?
 - c. Erforschung von Stereotypisierung und Vorurteilen in der Kommunikation: Inwiefern wirken sich Stereotype und Vorurteile auf die Kommunikation aus und wie können diese überwunden werden?
5. Korrelationsanalysen:
 - a. Herstellung von Zusammenhängen zwischen einzelnen stimmlichen und artikulatorischen Kriterien und der Gesamtwirkung des Vortrags: Welche stimmlichen Merkmale sind beispielsweise besonders für die Verständlichkeit, Kompetenzwirkung etc. relevant?
 - b. Untersuchung nonverbaler Kommunikationsmittel und deren Einfluss auf den Vortragsstil: Welche Rolle spielen Gestik, Mimik, Blickkontakt und Körperhaltung für die rhetorische Sprechwirkung?

Die Covid-19-Pandemie hat die Implementierung digitaler Technologien in der Hochschulbildung erheblich beschleunigt. Der zunehmende Einsatz von Online-Lehrformaten bietet sowohl Chancen als auch Herausforderungen. Wendt und Neumann (2024, S. 13-14) argumentieren, dass nicht das Medium selbst, sondern die Lernumgebung und die Qualität der Inhalte ausschlaggebend für den Lernerfolg sind. Neben den bekannten Vorteilen wie

Innovation, Zeiteffizienz, Ortsunabhängigkeit und Nachhaltigkeit⁵¹ identifizierte dieses Pilotprojekt neue Aspekte der digitalen Kommunikation.

Bemerkenswert war die unerwartet positive Bewertung des Publikumskontakts im Online-Format. Dies deutet darauf hin, dass Studierende sich erfolgreich an die digitale Lehr- und Lernumgebung angepasst haben. Sie entwickelten spezifische Techniken wie die Aufrechterhaltung virtuellen Blickkontakts und die Fähigkeit, auch online ein Gefühl der direkten Ansprache zu vermitteln. Diese Kompetenzen sind im Kontext der zunehmenden Digitalisierung von erheblicher Bedeutung.

Bionotes: Ulrike A. Kaunzner is a Professor of German Language and Linguistics at the University of Modena and Reggio Emilia. She earned her Ph.D. in German Linguistics from the University of Duisburg and enhanced her expertise through advanced studies in Communication Psychology and Intercultural Communication in both Germany and the United States. From 1991 to 2004, Prof. Kaunzner served as a CEL/Lecturer at the University of Bologna. Since 1996, she has been a valued member of the examination board for the Master of Speech Communication and Rhetoric program at the University of Regensburg. Her research and teaching interests encompass rhetoric and conversation techniques, speech and voice impact research, German phonetics and phonology, audiovisual translation, advertising language, and intercultural communication.

Lisa Bardeck holds degrees in Teacher's Training Study Program (Italian & German) from Paris Lodron University of Salzburg and in Foreign Languages and Literatures from the University of Florence (Università degli Studi di Firenze). She is currently a language expert (CEL) at the University of Modena and Reggio Emilia (Unimore) and collaborates with Unicollège SSML in Florence as a lecturer. At Unimore, she has contributed to the CAP research project. She is interested in foreign language didactics, contrastive linguistics (Italian-German), and translation studies.

Authors' addresses: ulrike.kaunzner@unimore.it; lisa.bardeck@unimore.it

⁵¹ In diesem Zusammenhang muss im Auge behalten werden, dass diese Vorteile nur dann wirklich zum Tragen kommen, wenn die jeweilige virtuelle Umgebung fachgerecht und didaktisch sinnvoll genutzt wird.

Literatur

- Allhoff D.-W. 1984, *Verständlichkeit gesprochener Sprache. Zum Stand der Forschung*, in „Sprechen“ 10/1984, S. 16-30.
- Allhoff D.-W. und Allhoff W. 2014, *Rhetorik & Kommunikation: Ein Lehr- und Übungsbuch*, Reinhardt, München/ Basel.
- Anderson L.W. und Krathwohl D.R. 2001, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, Longman, New York.
- Apel H., Méndez J. und von Laguna K. 2015, „Ihre Präsentation hat uns gut gefallen, aber...“. *Rhetorische Beurteilungskriterien auf dem Prüfstand: Fundierung anhand eines Anwendungsbeispiels*, in Teuchert B. (Hrsg.), *Mündliche Kommunikation lehren und lernen. Facetten der Rhetorik in Schule und Beruf*, (Sprache und Sprechen 48), Schneider Verlag Hohengehren, Baltmannsweiler, S. 107-117.
- Bose I. und Hannken-Illjes K. 2017, *Gesprächsrhetorik in der Sprechwissenschaft*, in Hess-Lüttich E. (Hrsg.), *Handbuch Gesprächsrhetorik*, De Gruyter, Berlin, S. 507-523.
- Büring D. 2006, *Intonation und Informationsstruktur*, in Blühdorn H., Breindl E. und Waßner U.H. (Hrsg.) *Text-Verstehen: Grammatik und darüber hinaus*, De Gruyter, Berlin, S. 144-163.
- Diadori P. 2015, *Le variabili nell'apprendimento della L2*, in Diadori P. (a cura di), *Insegnare italiano a stranieri*, Mondadori Education, Milano, S. 2-12.
- Draxler C. 2021, *OCTRA – Transcribing Speech in your Browser. A Tutorial*. – Interne Präsentation an der Unimore (8.10.2021)
- Draxler C. 2022, *Automatic Transcription of Spoken Language Using Publicly Available Web Services*, in Saturno J. e Spreafico L. (a cura di), *Fare linguistica applicata con le digital humanities*, AIItLA, Milano, S. 27-47.
- Erler M., und Tornau C. (Hrsg.) 2019, *Handbuch Antike Rhetorik*, Bd. 1, (Handbücher Rhetorik), De Gruyter, Berlin.
- Fišer D. and Witt A. (eds.) 2022, *CLARIN. The Infrastructure for Language Resources*, De Gruyter, Berlin.
- Fix U., Gardt A. und Knappe J. (Hrsg.) 2008, *Rhetorik und Stilistik / Rhetoric and Stylistics*, (HSK) 31, De Gruyter, Berlin.
- Forster R. 1997, *Mündliche Kommunikation in Deutsch als Fremdsprache: Gespräch und Rede*, Röhrig Universitätsverlag, St. Ingbert.
- Forster R. 2004, *Sprecherziehung im Unterricht „Deutsch als Fremdsprache“*, in Pabst-Weinschenk, M. (Hrsg.), *Grundlagen der Sprechwissenschaft und Sprecherziehung*, Reinhardt Verlag, München/ Basel, S. 324-28.
- Geißner H. 1982, *Sprecherziehung: Didaktik und Methodik der mündlichen Kommunikation*, Scriptor, Frankfurt am Main.
- Geißner H. 2000, *Kommunikationspädagogik. Transformationen der „Sprech“-Erziehung. Sprechen und Verstehen 1*, Bd. 17, Roehrig-Universitätsverlag, St. Ingbert.
- Groeben N. 1976, *Verstehen, Behalten, Interesse. Übereinstimmende Antworten und kontroverse Fragen zur Beziehung von Textstruktur, Textverständnis und Lerneffekt*, in „Unterrichtswissenschaft“ 2, S. 128-42.
- Grzeszczakowska-Pawlikowska B. 2016, *Rhetorische Kompetenz bei polnischen Studierenden in der Fremdsprache Deutsch im universitären Kontext*, in „Acta Universitatis Wratislaviensis Studia Linguistica XXXV“ 374, S. 425-438.
- Kaunzner U.A. 2018, *Das klingt sympathisch! Selbst- und Fremdbild in der Sprechwirkung*

- des italienischen Akzents*, in Vogt, B. (Hrsg.), *Deutsch als gesprochene (Fremd-) Sprache in Forschung und Lehre / Tedesco come lingua (straniera) parlata nella ricerca e nella didattica* 1, EUT, Trieste, S. 139-155.
- Kaunzner U.A. 2015, *Die Verständlichkeit vorgelesener Texte durch nichtdeutsche Sprecher. Zur Sprechwirkung bei muttersprachlichen und nicht-muttersprachlichen Rezipienten*, in Teuchert B. (Hrsg.), *Aktuelle Forschungstendenzen in der Sprechwissenschaft. Normen, Werte, Anwendung*, (Sprache und Sprechen, 48), Schneider Verlag Hohengehren, Baltmannsweiler, S. 65-81.
- Kisler T., Schiel F. and Sloetjes H. 2012, *Signal Processing via Web Services: the Use Case WebMAUS*, talk presented at Digital Humanities Conference 2012, Hamburg.
- Klein W. und Dimroth C. 2003, *Der ungesteuerte Zweitspracherwerb Erwachsener: Ein Überblick über den Forschungsstand*, in Maas U. und Mehlem U. (Hrsg.), in *Qualitätsanforderungen für die Sprachförderung im Rahmen der Integration von Zuwanderern*, (IMIS-Beiträge, 21), IMIS Osnabrück, S. 127-161.
- Langer I., Schulz von Thun F. und Tausch R. 1975, *Sich verständlich ausdrücken*, E. Reinhardt, München / Basel.
- Lapteva O. 2011, *Speaker Perception and Recognition. An Integrative Framework for Computational Speech Processing*, Kassel UP GmbH, Kassel.
- Michler C. 2015, *Einführung in die Didaktik der romanischen Sprachen und Literaturen*, University of Bamberg Press, Bamberg.
- Pömp J. und Draxler C. 2019, *Octra. Web-basierter Editor für die orthographische Transkription*, Zenodo, <https://zenodo.org/records/12771950> [27.07.2024].
- Selinker L. 1972, *Interlanguage*. *IRAL – International Review of Applied Linguistics*, in „Language Teaching“ 10 [3], S. 209–231.
- Stock E. 1991, *Grundfragen der Sprechwirkungsforschung*, in Krech, E.-M., Richter G. Stock E. und Suttner (Hrsg.), *J. Sprechwirkung: Grundfragen, Methoden und Ergebnisse ihrer Erforschung*, Akademie Verlag, Berlin, S. 9-27.
- Till D. 2015, *Rhetorik und Sprechkunst: Analyse und Anwendung*, in Teuchert, B. (Hrsg.), *Aktuelle Forschungstendenzen in der Sprechwissenschaft. Normen, Werte, Anwendung*, (Sprache und Sprechen, 48), Schneider Verlag Hohengehren, Baltmannsweiler, S. 1-15.
- Trim J.L.M., North B., Coste D., Sheils J., Quetz J., Schneider G. und Conseil de l'Europe 2001, *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*, Langenscheidt, Berlin.
- Ueding G. 2011, *Grundriss der Rhetorik*, Metzler, Stuttgart.
- Wendt C. und Neumann A. 2024, *Entwicklung eines Lehrkonzepts zur Implementierung von Medien für einen digital-gestützten Deutschunterricht im berufsbildenden Lehramt*, in Ahlers M., Besser M., Herzog C., Kuhl P., *Digitales Lehren und Lernen im Fachunterricht. Aktuelle Entwicklungen, Gegenstände und Prozesse*, Beltz Juventa, Lüneburg, S. 10-33.
- Wisniewski K. 2018, *Sprache und Studienerfolg von Bildungsausländerinnen und -ausländern: Eine Längsschnittstudie an den Universitäten Leipzig und Würzburg*, in „Info DaF“ 45 [4], S. 573-579.

Software und Internetseiten:

Avidemux (Software), <https://avidemux.sourceforge.net/> [14.05.2024]

Audacity (Software), <https://www.audacityteam.org/> [14.05.2024]

- BAS CLARIN Repository, <https://clarin.phonetik.uni-muenchen.de/BASRepository/>
[10.06.2024]
- LACOM Unimore, <https://www.unimore.it/it/didattica/corsi-di-studio/languages-communication-international-enterprises-and-organizations> [11.08.2024]
- LCE Unimore, <https://www.unimore.it/it/didattica/corsi-di-studio/lingue-e-culture-europee>
[11.08.2024]
- OCTRA, <https://clarin.phonetik.uni-muenchen.de/apps/octra/octra/login> [24.07.2024]
- Online-Fragebogen Pilotprojekt CAP, https://docs.google.com/forms/d/1oYCRAuv-Pet0t726iWh5iBc23xlkGbEAoIFGMrZbGWU/edit?usp=forms_home&ths=true
[15.05.2024]
- WebMAUS, <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic> [24.06.2024]

ANALYSE DER STIMMLICHEN LEISTUNGSFÄHIGKEIT BEI LEHRAMTSSTUDIERENDEN Ein Vergleich zweier digital gestützter multiparametrischer Verfahren

CHRISTIAN GEGNER
UNIVERSITÄT REGENSBURG

Abstract – Teachers belong to the group of professional voice users. The demands on vocal performance are correspondingly high. Since teachers are rarely prepared for their voice-intensive profession, the prevalence of voice disorders is particularly high in this professional group. In order to improve the training conditions for student teachers, the possibility of a digitally supported voice screening examination and the comparison of two multi-parametric methods are presented in this article: the procedure, the collected parameters and the results of a sample of about 200 students at the University of Regensburg, Germany. The conclusion is an outlook on the development of an app for voice hygiene and prophylaxis for people in teaching professions.

Keywords: teachers' voice; voice disorders among teachers; voice screening examination; voice hygiene and prophylaxis.

There is an emerging universal agreement among researchers within this area that preventive measures for voice disorders should be taken. Potential preventive strategies, such as voice screening, vocal health education, and voice training, may be required for voice disorders among the professional voice user population. These potential preventive strategies could be developed as prevention tools for professional voice users.

(Hazlett *et al.* "Review of the Impact of Voice Training on the Vocal Quality of Professional Voice Users", 2011, p. 189).

1. Der Lehrberuf – ein professioneller Sprechberuf

Der Lehrberuf ist international ein stimm- und sprechintensiver Beruf. Lehrkräfte unterrichten, führen Gespräche mit Schüler:innen, Kolleg:innen

und Eltern – viele Stunden am Tag und pro Woche. Ihre Stimme ist dabei das wichtigste Werkzeug, um Inhalte zu vermitteln und mit anderen Personen in den unterschiedlichsten Kontexten zu kommunizieren. Aus medizinischer Sicht gehören Lehrkräfte der Berufsgruppe der „professional voice users“ an (Schneider-Stickler, Bigenzahn 2013). Dies bedeutet, dass Lehrer:innen auf ihre Stimme zur Ausübung ihres Berufs angewiesen sind und zugleich der Lehrberuf und seine Rahmenbedingungen eine besondere Belastung für die Stimmgesundheit der jeweiligen Lehrkraft darstellen können. Das Risiko, im Verlauf der beruflichen Karriere eine Stimmstörung zu entwickeln, ist hoch. So zeigt eine Vielzahl von Studien, dass signifikant mehr Lehrkräfte im Vergleich zu Angehörigen anderer Berufsgruppen an Erkrankungen der Stimme leiden (Lehrkräfte: 11% bis 81% vs. Nicht-Lehrkräfte: 1,0% bis 36,1%) (Van Houtte *et al.* 2001). Martins *et al.* 2014 gehen aufgrund ihres Reviews zu Untersuchungen aus den Jahren 1997 bis 2013 bei Lehrkräften von einer zwei- bis dreimal höheren Prävalenz von Stimmstörungen als in der Allgemeinbevölkerung aus. Zudem scheinen Lehrerinnen im Vergleich zu ihren männlichen Kollegen ein 1,6-fach gesteigertes Risiko zu besitzen (Byeon 2019). Die Folgen von Stimmerkrankungen bei Lehrkräften sind weitreichend: Sie können nicht nur die berufliche Leistungsfähigkeit einschränken, sondern sogar den beruflichen Einsatz unmöglich machen (Schneider-Stickler, Bigenzahn 2013) oder der Grund für einen Berufswechsel sein (Behlau *et al.* 2012). Hierdurch entstehen dem Schul- und Gesundheitssystem enorme Kosten (Richter, Echternach 2010). Ebenso ergeben sich negative Auswirkungen auf den Lehr-Lernkontext, wie z. B. eine verminderte Verstehens- oder Behaltensleistung oder geringere Motivation bei Schüler:innen (Voigt-Zimmermann 2017; Murgia *et al.* 2022). Negative Auswirkungen auf das private Kommunikationsverhalten der betroffenen Lehrkraft sind ebenfalls nicht ausgeschlossen (Hwa Chen *et al.* 2010).

In Anbetracht dieser Auswirkungen und vor dem Hintergrund des europaweiten Mangels an Lehrer:innen (Beise 2023) besteht hier ein deutlicher Handlungsbedarf, wobei der Prävention von Stimmerkrankungen gegenüber der Nachsorge der Vorrang eingeräumt werden sollte.

2. Über die Notwendigkeit stimmpräventiver Maßnahmen in der Lehrkräftebildung

Als Hauptrisikofaktoren für die Stimmgesundheit einer Lehrperson gelten beispielsweise: Geschlecht, Erkrankungen der oberen Atemwege, Konsum koffeinhaltiger Getränke, lautes Sprechen, Anzahl der Unterrichtsstunden pro Woche und Resignationserfahrungen aufgrund stimmlicher Probleme (Byeon 2019). Ebenso werden ein hoher Umgebungslärm (Martins *et al.* 2014), die Stressbelastung (Gasull *et al.* 2010; Vertanen-Greis *et al.* 2020) oder auch eine

mangelnde Stimmkonstitution (Schneider *et al.* 2004) mit Stimmproblemen bei Lehrkräften assoziiert. Auf Basis eigener Untersuchungen weist Meuret (2017) zudem darauf hin, dass das Risiko, an einer Stimmstörung zu erkranken, um das 1,6-fache steigt, wenn während des Studiums keine präventiven Maßnahmen wie Stimm- und Sprecherziehung stattfinden.

Somit kann der Prävention eine entscheidende Bedeutung für die Förderung der Stimmgesundheit bei Lehrkräften beigemessen werden. Grundsätzlich werden drei Arten der Prävention im Hinblick auf den Zeitpunkt der Intervention unterschieden:

- (1) Die **Primärprävention** konzentriert sich auf vorbeugende Maßnahmen, um Gesundheitsrisiken zu verringern, bevor Symptome auftreten: z. B. Angebote während des Studiums.
- (2) Die **Sekundärprävention** zielt auf die Früherkennung und Behandlung von Krankheiten ab, um deren Fortschreiten zu verhindern.
- (3) Die **Tertiärprävention** versucht Krankheitsfolgen oder auch Rückfälle zu mindern und wird teilweise mit der medizinischen Rehabilitation gleichgesetzt (Rittich 2018).

Vor dem Hintergrund der Primärprävention von Stimmstörungen ist es bedeutsam, dass im Unterschied zu Angehörigen der Hochleistungsstimmerufe, z. B. Sänger:innen und Schauspieler:innen, künftige Lehrkräfte kaum oder in vielen Fällen nur unzureichend auf ihren professionellen Sprechberuf vorbereitet werden (Eidenmüller 2016). In Deutschland sind nur in wenigen Bundesländern verpflichtende Angebote wie Stimm- und Sprecherziehung in die Lehramtsstudiengänge integriert, wobei auch der Umfang in Abhängigkeit von der jeweiligen Hochschule variiert (Lange, Appel 2014; Hukelmann 2013; Sick 2019). In Italien erhalten Lehrkräfte keinerlei stimmliche Präventionsangebote (Angelillo *et al.* 2009).

Im Übergang von der Primär- zur Sekundärprävention existieren in Deutschland inzwischen immer mehr – meist fakultative – Beratungsangebote an Hochschulen, um stimmauffällige Studierende zu identifizieren und Betroffenen rechtzeitig Hilfen anbieten zu können, damit sie die nötigen stimmlichen und sprecherischen Kompetenzen für ihren späteren Beruf als „professional voice users“ erwerben können: z. B. das Stimmscreening-Angebot an der Universität Regensburg, an der Universität Aachen, an der Pädagogischen Hochschule Weingarten oder der Stimmcheck an der Universität Halle-Wittenberg (Christmann *et al.* 2024; Gegner, Schilcher 2022; Gegner *et al.* 2024; Hillegeist, Thomas 2024). In diesem Zusammenhang stellt sich jedoch die Frage, wie derartige Beratungsformate konzipiert sind und welche Untersuchungsmethoden hierbei zur Anwendung kommen. Aus diesem Grund wird im Folgenden die Möglichkeit einer digital gestützten Analyse der stimmlichen Leistungsfähigkeit von Lehramtsstudierenden vorgestellt. Hierbei werden der Ablauf, die verwendete Software und die erhobenen Parameter

einer Stimmscreening-Untersuchung dargestellt sowie die Ergebnisse der stimmlichen Leistung von ca. 200 Studierenden vorgestellt, die das Angebot seit dem Wintersemester 2010/11 in Regensburg in Anspruch genommen haben. In diesem Zusammenhang werden zwei multiparametrische Verfahren zur Bestimmung der Stimmqualität und -quantität verglichen: der „Dysphonia Severity Index“ (Wuyts *et al.* 1997; 2000) und der Dysphonie-Index nach Friedrich (1998). Den Abschluss bildet ein Ausblick auf das digitale Präventionsprojekt „ReSt – Regensburger Stimmtraining: Eine App zu Stimmgesundheit und -prophylaxe“. Die App richtet sich an bereits tätige Lehrkräfte.

3. Das Stimmscreening – eine digital gestützte Analyse der stimmlichen Leistungsfähigkeit

3.1. Hintergrund: Stimmdiagnostik

Für die Untersuchung und auch Beurteilung der menschlichen Stimme und deren Funktion existieren eine Vielzahl von Methoden. Diese leiten sich teils aus der Profession der Untersuchenden (z. B. Medizin, Therapie, Pädagogik) her oder dienen dazu, einzelne Funktionsbereiche der Stimmproduktion (z. B. Haltung und Tonus, Atmung, Phonation, Artikulation) zu untersuchen.

Aus stimmärztlicher Sicht kann das Basisprotokoll der „European Laryngological Society (ELS)“ als Minimalkonsens gelten, welches subjektive und objektive Methoden mit dem Ziel einer multidimensionalen Bewertung der Stimme (Schneider-Stickler, Biegenzahn 2013) kombiniert. Es beinhaltet folgende Elemente:

- Perzeption,
- Videostroboskopie¹,
- akustische Analysen²,
- aerodynamische Messungen³,
- subjektive Selbstevaluation durch die/den Patient:in (Dejonckere *et al.* 2001).

¹ Stroboskopie ist ein Visualisierungsverfahren, mit welchem der „Schwingungsablauf der Stimmlippen“ untersucht wird (Seidner, Eysholdt 2015, S. 113).

² „Instrumentelle oder objektive akustische Analysen sind ein Sammelbegriff für computerbasierte Messungen der akustischen Eigenschaften einer Stimme“ (Brockmann-Bauser, Bohlender 2014, S. 49).

³ Aerodynamische Messungen liefern „Informationen über die Leistungsfähigkeit des Glottisgenerators bei der Umsetzung von Luftdruck in Stimmenschall“ (Seidner, Eysholdt 2015, S. 107). Hierzu zählt z. B. die Messung der maximalen Phonationszeit (MPT).

Als Minimalkonsens orientiert sich das Protokoll an den „Anforderungen der evidenzbasierten Medizin“ (Richter, Echternach 2007) und kann um zusätzliche Untersuchungsmethoden bei Bedarf ergänzt werden (Brockmann-Bausser, Bohlender 2014). Allerdings werden Stroboskopien in Deutschland ausschließlich von ärztlichem Personal durchgeführt. Deshalb schlagen Phoniater:innen für eine Stimmleistungsbewertung an einer pädagogischen Einrichtung wie einer Hochschule folgende Bestandteile vor: die auditive Stimmklangbeurteilung (= Perzeption), die Stimmfeldmessung (= akustische Analysen) und die Beurteilung von Artikulation und Myofunktion⁴ (= Perzeption) (Schneider-Stickler 2017). Die Leistungsbewertung erfordert somit kein ärztliches, aber geschultes Personal und es kann zusätzlich auf digitale Analysetechnik zurückgegriffen werden. Für akustische Analysen des Stimmschalls – im Besonderen die Stimmfeld-/ Stimm- umfangsprofilmessung – werden von unterschiedlichen Anbietern international Hard- und Softwarepakete bereitgestellt, die nach Analysemöglichkeiten und Umfang variieren. Derzeit existieren folgende Angebote: z. B. vom Stimminstitut VOCE (VidiVoice), von XION (DiVAS) oder von WEVOSYS (lingWAVES). Eine weitere Möglichkeit zur digitalen Analyse akustischer Stimmparameter bietet das Praat-Stimmprofil – jedoch ohne Stimmfeld-/Stimmumfangsprofil. Gerade die Stimmfeld-/ Stimmumfangsprofilmessung ermöglicht jedoch die Überprüfung der stimmlichen Leistungsfähigkeit sowie intra- und interindividuelle Vergleiche (Schneider-Stickler, Bigenzahn 2013; Schultz-Coulon 1990). Stimmumfangsprofile oder Stimmfelder „beruhen auf einer quantitativen computergestützten Messung des Lautstärken- und Tonhöhenumfangs einer Stimme beim Sprechen (Sprechstimmprofil) und Singen (Singstimmprofil)“ (Brockmann-Bausser, Bohlender 2014, S. 49). Für die Erstellung der Leistungsprofile werden bei der Untersuchung der Singstimme der minimale und maximale Stimmschalldruck gehaltener Töne in Bezug zum Tonhöhenumfang verzeichnet (Nawka *et al.* 2006) und bei der Untersuchung der Sprechstimme „Schalldruckpegel und Frequenzen beim Sprechen in unterschiedlicher Lautstärke bestimmt“ (Eichel 2015, S. 216). Je nach Software können noch andere akustische Analysen des Stimmschalls vorgenommen werden, welche die Irregularität (z. B. Jitter, Shimmer) des Stimmsignals und auch additives Rauschen (z. B. Glottal-to-Noise-Excitation-Ratio (GNE)) anhand kurzer Stimmproben analysieren (Seidner, Eysholdt 2015). Durch die Integration aerodynamischer Messmethoden, wie z. B. die maximale Phonationszeit, können zudem multiparametrische Verfahren zur Beschreibung der Stimmquantität und -qualität (Friedrich 2013) wie der „Dysphonia Severity Index“ (Wuyts *et al.* 1997; 2000) und durch die Kombination mit perzeptiven Untersuchungsmethoden und der

⁴ Hierunter werden z. B. „Störungen der Muskelfunktion, des Muskeltonus oder der harmonischen Bewegungsabläufe“ (Hahn 2009, S. 327) im Mund- und Gesichtsbereich verstanden.

Selbstevaluation durch die/den Klient:in der Dysphonie-Index nach Friedrich (1998) bestimmt werden.

Wie die dargestellten Untersuchungsmethoden für eine Stimmleistungsbewertung von Studierenden an einer Hochschule genutzt werden können, wird in den nächsten Abschnitten exemplarisch am Ablauf einer Screening-Untersuchung, wie sie derzeit an der Universität Regensburg angeboten wird, vorgestellt. Das Stimmscreening beinhaltet die Selbstevaluation durch Klient:innen, perzeptive, instrumentell-akustische sowie aerodynamische Messmethoden.

3.2. Stimmscreening-Ablauf und -Referenzwerte

An der Universität Regensburg wird das Stimmscreening seit dem Wintersemester 2010/11 angeboten. Hierbei handelt es sich um ein fakultatives Beratungsangebot, welches sich in der Anfangsphase vor allem an Lehramtsstudierende mit dem Unterrichtsfach Deutsch wendete. Seit 2019 steht das Angebot im Rahmen des neu gegründeten „Communication and Voice Center for Teachers“ (CoVoC-T)⁵ allen Studierenden und Lehrenden der Universität Regensburg zur Verfügung. Ein zentrales Anliegen des CoVoC-T ist es, in einem deutschlandweiten Austausch mit anderen Universitäten und Pädagogischen Hochschulen („Netzwerk Stimmgesundheit“), eine Standardisierung hinsichtlich der Inhalte, Referenzwerte und des Ablaufs weiter voranzutreiben und eine Sensibilisierung für die Bedeutsamkeit von Stimme und Sprechen im Lehrberuf in der Bevölkerung zu erreichen. An der Universität Regensburg wird das Screening von Vertreter:innen des Fachbereichs Sprechwissenschaft/Sprecherziehung durchgeführt und dauert im Durchschnitt ca. 45min.

3.2.1. Bestandteile

In seiner aktuellen Version besteht das Stimmscreening aus drei zentralen Phasen: (1) Anamnese, (2) Erhebung eines Stimm- und Sprechstatus/Analyse der stimmlichen Leistungsfähigkeit, (3) Feedback/Aufklärung/Rat. Die erste und zweite Phase besteht wiederum aus mehreren Teiluntersuchungen. Für die Analysen und das Klient:innenmanagement wird die Software lingWAVES (Version 3.2.7) der Firma WEVOSYS verwendet. Die entsprechenden Module der Software werden in Klammern hinter der jeweiligen Teiluntersuchung aufgeführt. Die Darstellung erfolgt nach Gegner *et al.* (2024):

⁵ Das CoVoC-T wurde im Rahmen der gemeinsamen „Qualitätsoffensive Lehrerbildung“ von Bund und Ländern mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01JA1812 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

- **Kurzanamnese** (Modul *Sprechstatus*)
- **Selbsteinschätzung durch die/den Klient:in** (Modul *VDI = Voice Diagnostic Index*)
- **Perzeptive Einschätzung von Körperhaltung, Atmung, Stimme und Sprechen** (Modul *Sprechstatus, Aufnahme*): Tonus (im Stehen); Atmung; Stimmklang; Stimmeinsatz, -absatz, -sitz; Myofunktion/Artikulation
- **Sprechstimmfeldmessung** (Modul *Stimmfeld VDC*): durchschnittliche Sprechstimmlage und der durchschnittliche Schalldruckpegel für normale, leise und laute Phonation sowie die Rufstimme; Sprechstimmfeldanstieg; Tonhöhen und Dynamikumfang der Sprechstimme; Schalldruckpegel der leisesten und lautesten Phonation
- **Singstimmfeldmessung** (Modul *Stimmfeld VDC*): Tonhöhen- und Dynamikumfang der Singstimme; Normsingstimmfeldabdeckung; Tonhöhendynamik bezogen auf bestimmte Lautstärkestufen; Lautstärkedynamik bezogen auf verschiedene Tonhöhenstufen
- **Akustische Analysen des Stimmschalls** (Modul *Stimmfeld VDC*): Irregularität, Rauschen, Gesamtgrad, Jitter, Shimmer, Glottal-to-Noise-Excitation-Ratio (GNE)
- **Aerodynamische Messung** (Modul *Stimmfeld VDC*): maximale Phonationszeit (MPT)
- **Multiparametrisches Verfahren** (Modul *Stimmfeld VDC*): *Dysphonia Severity Index* (DSI)
- **Rhetorische Analyse** (Modul *Aufnahme*)

Der Screening-Ablauf ist standardisiert und jede Beratung findet individuell mit einer/einem Klient:in statt. Bei den akustischen Aufnahmen werden die Vorgaben von Nawka *et al.* (2006) und Schneider-Stickler (2016) eingehalten. In der Abschlussphase erhält jede:r Klient:in ein ausführliches Feedback über seine/ihre stimmliche Qualität und Leistungsfähigkeit aufgrund der Selbsteinschätzung, der perzeptiven und akustischen Untersuchung. Auf Basis der einzelnen Ergebnisse und des Gesamteindrucks wird nach Ausmaß der stimmlichen Auffälligkeit betroffenen Studierenden geraten, eine:n Fachärzt:in für Stimm-, Sprech-, Sprachstörungen zu konsultieren oder Kurse im Fachbereich Sprecherziehung zu besuchen.

In der Folge werden der Untersuchungsablauf und die gängigen Referenzwerte in Anlehnung an Gegner (2020) und Gegner *et al.* (2024) näher beschrieben.

3.2.2. Untersuchungsablauf und Referenzwerte

Zu Beginn der Untersuchung werden im Rahmen einer Kurzanamnese die Gründe der Klient:innen für die Konsultation, allgemeine

Unverträglichkeiten/Allergien (z. B. Hausstaub, Tierhaare), Angaben zu vergangenen oder aktuellen stimmärztlichen Diagnosen und logopädischen Therapien sowie zu allgemeinen Stimm- und Sprechanstrengungen erfragt. Im Anschluss erfolgt eine kurze Schilderung des Ablaufs. Den Abschluss der Anamnese phase bildet die Selbsteinschätzung der stimmlichen Situation mittels des in der Software integrierten „Voice Diagnostic Index“ (VDI)/Stimmstörungsindex (SSI)/Voice-Handicap-Index mit 12 Items (VHI-12; Nawka, Gonnermann 2003). Das Ergebnis wird von der Software ausgewertet und mit einem Ampelsystem visualisiert: grün = normal; hellgrün = leicht gestört; hellrot = mittelgradig gestört; rot = hochgradig gestört).

Die perzeptive Einschätzung von Körperhaltung, Atmung, Stimme und Sprechen durch die Untersuchungsleitung erfolgt mittels des Standardlesetextes „Nordwind und Sonne“ und einiger Testsätze, welche stehend von der/dem jeweiligen Klient:in angesprochen werden. Als unauffällig gelten Personen mit eutoner Körperhaltung, einer kostoabdominalen Atmung, einem Stimmklang nach der RBH-Klassifikation⁶ mit $R=\max 1$, $B=0$, $H=0$, physiologischen Stimmeinsätzen und -absätzen sowie einem vorderen bis mittleren Stimmsitz und einer unauffälligen Artikulation und Myofunktion ohne Lautbildungsfehler. Dies ist besonders auch wegen des ausgeprägten Modellcharakters von Lehrkräften im Primarbereich relevant.

Für die Messung der Sprechstimme wird der/die Klient:in angehalten, die Zahlenreihen von 21 bis 40 wie beim Zählen vorbeifahrender Autos in folgenden Lautstärken zu sprechen: normale Sprechlautstärke, so leise wie möglich (ohne zu flüstern), Vortragslautstärke. Die Messung der Rufstimmleistung erfolgt anhand des Testsatzes „Hey Anton, komm mal rüber!“. Als Referenz sollen die Klient:innen bei der leisen Sprechstimme einen Durchschnittswert von ≤ 50 dB (A) nicht überschreiten und für die Rufstimme einen Wert > 90 dB (A) erreichen (Schneider-Stickler 2016). Die durchschnittliche Lage der mittleren Sprechstimme (normale Lautstärke) wird in Bezug zur unteren Grenze des Stimmumfangs (Singstimme) gesetzt. Die Lage sollte 3 – 8 Halbtöne (HT) oberhalb der unteren Grenze des Stimmumfangs nicht überschreiten (Nawka, Wirth 2008). Um zu überprüfen, ob keine unverhältnismäßige Tonhöhenzunahme im Vergleich zur erreichten Lautstärke vorliegt, wird von der Software noch der Sprechstimmfeldanstieg bestimmt. Als unauffällig gilt ein Anstieg zwischen 0,3 und 0,4 Halbtönen pro Dezibel (Nawka, Wirth 2008).

Für die Analyse der Singstimme werden die Klient:innen instruiert, gehaltene Töne auf /a:/ von der Mitte bis zum unteren Ende des Stimmumfangs und im Anschluss bis zum oberen Ende des Stimmumfangs (ohne ästhetischen Anspruch) zu phonieren. Die Messung des Tonhöhenumfangs für die leise

⁶ R = Rauheit, B = Behauchtheit, H = Heiserkeit (übergeordnete Kategorie); 0 = nicht vorhanden, 1 = geringgradig gestört, 2 = mittelgradig gestört, 3 = hochgradig gestört (Nawka, Anders 1996).

Singstimme erfolgt vor Messung des Tonhöhenumfangs für die laute Singstimme. Analog zur Sprechstimme sollen auch hier der minimale Schalldruckpegel beim leisen Singen ≤ 50 dB (A) und der maximale Schalldruckpegel beim lauten Singen > 90 dB (A) betragen (Schneider-Stickler 2016; Schneider-Stickler, Bigenzahn 2013) und in mindestens einer Frequenz eine Stimmdynamik zwischen lautem und leisem Singen von > 35 dB (A) besitzen (Eichel 2015). Als unauffällig gilt ein Tonhöhenumfang der leisen und lauten Singstimme von je > 24 HT (Friedrich 1998; Scheider-Stickler, Bigenzahn 2013).

Abbildung 1 zeigt ein Stimmumfangsprofil mit der Software lingWAVES und Werten einer stimmlich unauffälligen Person. Die blauen Felder für die normale (N), leise (S) und laute Sprechstimme (L) befinden sich im unteren Drittel des gesamten Stimmumfangs (Kontur im Hintergrund). Die Rufstimme (Δ) markiert „den frequenzmäßig höchsten sowie lautesten Bereich“ des Brustregisters (Hacki 1999, S. 815).

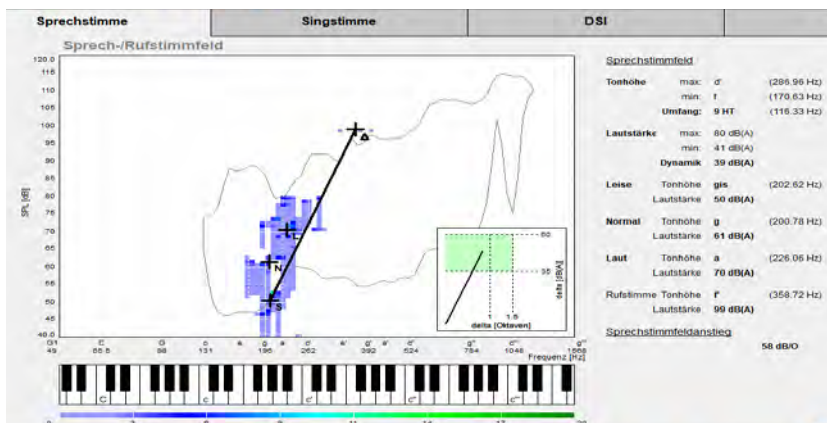


Abbildung 1

Stimmumfangsprofil einer stimmlich unauffälligen Person mit der Software lingWAVES.

Abbildung 2 zeigt im Vergleich ein Stimmumfangsprofil einer stimmlich auffälligen Person.

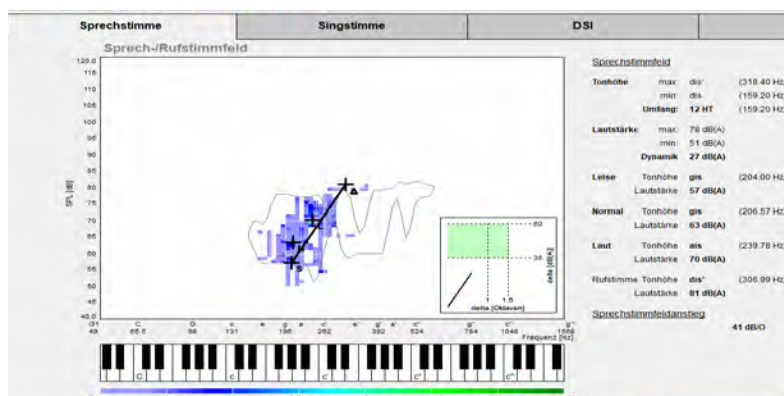


Abbildung 2

Stimmumfangsprofil einer stimmlich auffälligen Person mit der Software lingWAVES.

Deutlich erkennbar sind die Einschränkungen des Umfangs der Singstimme (Kontur im Hintergrund) sowie die Über- bzw. Unterschreitung der Referenzwerte für die leise Sprechstimme (S) und der Rufstimme (Δ) und die damit einhergehende mangelnden Stimmdynamik.

Die weiterführenden akustischen Analysen des Stimmschalls erfolgen anhand des Testwortes „Hase“, in welchem das /a:/ über ca. 3 Sekunden gehalten wird. Von der Software werden die Irregularitätsmaße Jitter (Schwankungen im Frequenzbereich) und Shimmer (Schwankungen der Amplitude) sowie als Maß für additives Rauschen die Glottal-to-Noise-Excitation-Ratio (GNE) bestimmt. Aufgrund softwarespezifischer Algorithmen werden – als objektive Anlehnung zur RBH-Bewertung – Irregularität, Rauschen und ein Störungsgrad ermittelt. Als unauffällig gelten Stimmen, die unter folgenden Grenzwerten liegen: Irregularität (< 1); Rauschen (< 1); Gesamtgrad (< 1); Jitter (max. 0,5%); Shimmer (max. 5,0%), GNE (max. 0,6). Die Grenzwerte sind in der Software hinterlegt.

Bei der aerodynamischen Messung wird die/der Klient:in gebeten, den Laut /a:/ so lange wie möglich bei normaler Sprechlautstärke und Tonhöhe zur Messung der maximalen Phonationszeit (MPT) zu phonieren. Die Messung wird dreimal wiederholt und der höchste Wert in Sekunden verzeichnet. Gesunde Stimmen sollten hier in der Lage sein, einen Wert > 15 s zu erreichen (Schneider-Stickler, Bigenzahn 2013).

Auf Basis einzelner Werte aus der Stimmfeldmessung/-umfangsprofilmessung, den akustischen Analysen des Stimmschalls und der maximalen Phonationszeit errechnet die Software den multiparametrischen „Dysphonia Severity Index“ (DSI) zur Beschreibung der Stimmqualität bzw. des Schweregrads einer Stimmstörung. In die Berechnung fließen die maximale Tonhaldedauer (MPT), die höchste erzielte Frequenz ($F0_{\max}$) der Singstimme, die niedrigste Intensität (SPL_{\min}) der Singstimme und der Jitter-Wert in Prozentangabe nach folgender Formel ein (Wuyts *et al.* 1997; 2000):

$$DSI = 0,13 \times MPT + 0,0053 \times F0_{\max} - 0,26SPL_{\min} - 1,18 \times \text{Jitter} (\%) + 12,4$$

Die DSI-Werte können nach dem lingWAVES Benutzerhandbuch (Version 3.2 – 2017-09, S. 54) zur Beschreibung der Stimmgüte in Intervalle eingeteilt und sechs Klassen zugeteilt werden. Tabelle 1 zeigt die Klassierung der Werte. Beispielsweise gelten Stimmen mit einem Wert $> 4,4$ als unauffällig (Klasse 5)⁷.

⁷ Nawka *et al.* (2006, S. 18) geben andere Werte an: $> 4,2$ = normale Stimme; 4,2 bis $> 1,8$ = leichte Dysphonie; 1,8 bis $> 1,2$ mittelgradige Dysphonie; $\leq 1,2$ = hochgradige Dysphonie.

Klasse	DSI-Bereich	Beschreibung
0	< -2,0	Anhaltende Dysphonie, keine Phonation möglich
1	-1,9 bis +0,3	Sporadische, periodische Phonation, aphonische Perioden
2	+0,4 bis +2,2	Patient kann Stimme produzieren, aber mit zahlreichen Dysphonieperioden
3	+2,3 bis +3,3	Wenige Perioden gestörter Phonation, leichte Dysphonie
4	+3,4 bis +4,3	Sporadische Dysphonie für kurze Momente
5	> 4,4	Keine Dysphonie

Tabelle 1

DSI-Klassen nach dem lingWAVES Voice Diagnostic Center (VDC) Benutzerhandbuch (Version 3.2 – 2017-09, S. 54).

Abbildung 3 zeigt die akustischen Analysen des Stimmchalls, die Messung der maximalen Phonationszeit sowie den DSI bei einer stimmunauffälligen Person mittels der Software lingWAVES.

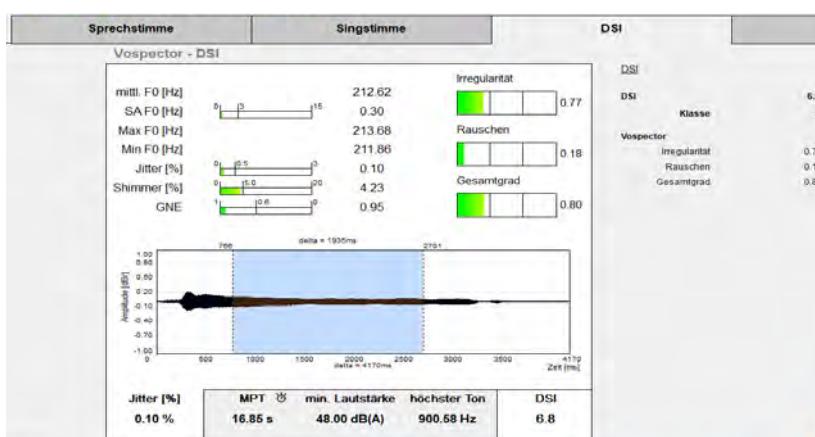


Abbildung 3

Akustische Analyse des Stimmchalls und Messung der maximalen Phonationszeit bei einer stimmlich unauffälligen Person mit der Software lingWAVES.

Im Vergleich hierzu verdeutlicht Abbildung 4 die Messergebnisse bei einer stimmlich auffälligen Person. Hier sind die Shimmer- und GNE-Werte sowie die Maße für Irregularität, Rauschen und der Störungsgesamtgrad deutlich erhöht. Die maximale Phonationszeit (MPT) liegt unterhalb von 10 Sekunden, was auf eine Pathologie hindeuten könnte (Schneider-Stickler, Bigenzahn 2013, S. 74). Der DSI-Wert beträgt 3,6 (Klasse 4), was nach dem lingWAVES Benutzerhandbuch (Version 3.2 – 2017-09, S. 54) als Anzeichen für eine „[s]poradische Dysphonie für kurze Momente“ gewertet werden kann.

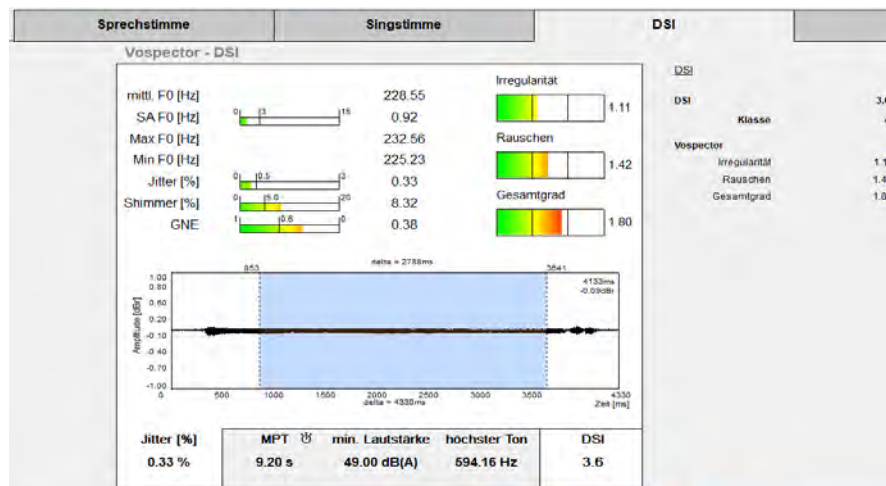


Abbildung 4

Akustische Analyse des Stimmschalls und Messung der maximalen Phonationszeit bei einer stimmlich auffälligen Person mit der Software lingWAVES.

Im weiteren Screening-Verlauf erfolgt die rhetorische Analyse. Hierfür werden die Klient:innen gebeten, einen frei gewählten Gegenstand (z. B. Inhaltsauszug aus dem letzten Referat) über eine Dauer von zwei Minuten wiederzugeben. Somit handelt es sich um einen Sprechbeitrag ohne schriftliche Grundlage, der für eine Abklärung der anfänglichen perzeptiven Einschätzung anhand des Standardlesetextes herangezogen wird. Beobungskriterien stellen generell die verbale, para- und extraverbale Ebene dar. Auf Basis dieses Eindrucks erhalten die Klient:innen auch eine kurze Rückmeldung zu ihren rederhetorischen Fähigkeiten.

Für einen ersten Überblick fasst Tabelle 2 die zentralen untersuchten und erfassten Parameter bei der Stimmscreening-Sprechstunde und die zu Grunde gelegten Referenzwerte (ohne rhetorische Analyse) noch einmal zusammen.

Untersuchungsbereich	Referenz
Selbsteinschätzung	
VDI (SSI/VHI-12)	Keine Störung
Perzeptive Untersuchung	
Körperhaltung im Stehen	euton
Atmung	kostoabdominal
RBH-Beurteilung:	$R_{\max}=1$; $B=0$; $H=0$
Stimmeinsätze	weich/fest
Stimmabsätze	weich/fest
Stimmansatz	vorne

Artikulation (mit Vermerk von Dialekt)	keine phonetische Störung, angemessene Kieferöffnung und Lippenbeteiligung
Instrumentelle Untersuchung (Sprech- und Singstimme)	
Sprechstimme: leisester Wert	≤ 50 dB
Rufstimme: lautester Wert	> 90 dB
Stimmdynamik zw. leiser Sprech- und Rufstimme	> 35 dB
Halbtonanstieg zw. Ø leiser Sprech- und Rufstimme	≤ 0,4 HT
Indifferente Sprechstimmlage (normale Lautstärke)	3 – 8 HT oberhalb der unteren Stimmgrenze
Anzahl der Halbtöne beim leisen Singen	> 24 HT
Anzahl der Halbtöne beim lauten Singen	> 24 HT
Minimaler Schalldruckpegel beim leisen Singen	≤ 50 dB
Maximaler Schalldruckpegel beim lauten Singen	> 90 dB
Differenz zwischen leisem und lautem Singen in mind. einer Frequenz	> 35 dB
Aerodynamische Messung	
Maximale Phonationszeit	> 15 s
Multiparametrisches Verfahren	
Dysphonia Severity Index (DSI)	> 4,4

Tabelle 2

Untersuchungsbereiche und Referenzwerte des Stimm screenings an der Universität Regensburg aus Gegner *et al.* (2024, S. 145).

Obwohl in der Untersuchungspraxis eine Gesamtberücksichtigung aller Bereiche für eine holistische Beschreibung der stimmlichen Situation stattfindet, bieten gerade multiparametrische Verfahren die Möglichkeit der Reduktion und Gewichtung mehrerer „als relevant definierte Stimmfunktionsparameter zu einem errechneten Index“ (Friedrich 2013, S. 78). Neben dem DSI existiert ein weiteres multiparametrisches Verfahren zur Abgrenzung von gesunden und erkrankten Stimmen sowie zur Beschreibung des Schweregrads einer Stimmstörung: der Dysphonie-Index nach Friedrich (1998). Im Gegensatz zum DSI kombiniert der Dysphonie-Index Ergebnisse der Selbstevaluation durch die/den Klient:in, der perzeptiven Einschätzung durch die untersuchende Person, der instrumentellen Untersuchung (Stimmumfang, Stimmdynamik) und der maximalen Phonationszeit. Alle Ergebnisse werden auf einer vierstufigen, diagnostischen Skala (0 = keine; 1 = geringgradig; 2 = mittelgradig; 3 = hochgradig) bewertet. Die Indexberechnung entspricht dem arithmetischen Mittel (Tab. 3).

Untersuchungsparameter	Skala			
	0	1	2	3
Heiserkeit (auditive Stimmklangbeurteilung, RBH)	keine	geringgradig	mittelgradig	hochgradig
Stimmumfang der Singstimme in HT	> 24	24 – 18	17 – 12	< 12
Stimmdynamik = Differenz zwischen leiser Sprechstimme und Rufstimme in dB (A)	> 45	45 – 35	34 – 25	< 25
Tonhaltedauer in Sekunden	> 15	15 – 11	10 – 7	< 7
Kommunikative Beeinträchtigung	keine	gering	mäßig	stark
Dysphonie-Index ($\Sigma/5$)				

Tabelle 3
Dysphonie-Index nach Friedrich 1998, S. 15.

Die kommunikative Stimmbeeinträchtigung wird mit folgender Skala erfragt:

Skala	Beschreibung
0	Keine kommunikative Beeinträchtigung
1	Keine Beeinträchtigung in der alltäglichen, sozialen Kommunikation; geringe Einschränkung bei verstärkter Stimmbelastung bzw. erhöhter Stimmanforderung
2	Geringe Beeinträchtigung in der alltäglichen, sozialen Kommunikation; starke Einschränkung bei verstärkter Stimmbelastung bzw. erhöhter Stimmanforderung, Stimme nicht belastbar
3	Starke Einschränkung auch in der alltäglichen Kommunikation, Sozialkontakte beeinträchtigt

Tabelle 4
Bewertung der kommunikativen Stimmbeeinträchtigung nach Friedrich 1998, S. 11.

Wie aus Tabelle 3 hervorgeht, können alle Werte aus der Stimmscreening-Untersuchung, wie sie an der Universität Regensburg durchgeführt wird, gewonnen werden. Lediglich die von Friedrich (1998) eingesetzte Selbsteinschätzung zur kommunikativen Beeinträchtigung aufgrund einer Stimmproblematik muss ersetzt werden.

Da in diesem Artikel besonders die digital gestützte Analyse der stimmlichen Leistungsfähigkeit, wie sie an einer Hochschule umgesetzt werden kann, im Fokus steht, werden in der Folge die Ergebnisse aus einer Teiluntersuchung von Gegner (2021) vorgestellt. Sie betreffen den Vergleich zwischen den multiparametrischen Verfahren des „Dysphonia Severity Index

(DSI)“ und des „Dysphonie-Index nach Friedrich (1998)“ zur Beurteilung der stimmlichen Leistungsfähigkeit einer studentischen Stichprobe.

4. Ergebnisse zur Stimmleistungsbewertung

Die Untersuchung von Gegner (2021) verfolgte das Ziel, den Prozentsatz an stimmlich auffälligen Lehramtsstudierenden mit dem Unterrichtsfach Deutsch zu identifizieren und die Fragen zu beantworten, über welche rederhetorischen und sprechkünstlerischen Fähigkeiten und Fertigkeiten die Studierenden verfügen. Überdies sollte überprüft werden, ob ein Zusammenhang zwischen der stimmlichen Leistungsfähigkeit der Studierenden und ihren rederhetorischen bzw. sprechkünstlerischen Kompetenzen besteht. Die Konzeptualisierung und Operationalisierung der Konstrukte zur sprechkünstlerischen und rederhetorischen Leistung sowie die Ergebnisse zu diesen Teiluntersuchungen werden bei Gegner (2021) und Gegner und Schilcher (2022) dargestellt.

4.1. Untersuchungsdesign

Zur Beantwortung der Forschungsfragen wurde der geschilderte Screeningablauf (vgl. 3.2.2) modifiziert. Zum einen wurde ein zusätzlicher Untersuchungsblock in den Ablauf integriert, der die Erhebung der sprechkünstlerischen Fähigkeiten anhand der Textvorlage „Erlkönig“ ermöglichte. Zum anderen wurde für die stimmliche Selbsteinschätzung der Studierenden ein eigener Fragebogen zu 13 symptomassoziierten Beschwerden konzipiert, deren Vorkommenshäufigkeit auf einer vierstufigen Skala eingeschätzt werden sollte („nie“, „selten“, „manchmal“, „häufig“): Heiserkeit, Stimmversagen, Stimmmüdigkeit, Schmerzen beim Sprechen, Räusperzwang, Druckgefühl, Engegefühl im Halsbereich, Kloßgefühl im Halsbereich, Kratzen im Halsbereich, Schluckzwang, Brennen im Halsbereich, Mundtrockenheit, Luftmangel beim Sprechen (zur Konstruktion: Gegner 2021, S. 93f.). Dieses Vorgehen erfolgte primär aus zwei Gründen. Zum einen sollten die Studierende nicht durch die Erfragung nach dem Vorhandensein einer Stimmstörung, wie im VHI, eingeschüchtert werden. Zum anderen ist davon auszugehen, dass sich ein stimmliches Selbstkonzept erst im Verlauf der Ausbildung und mit dem Erleben stimmlicher Belastungssituationen etabliert und somit bei Studienanfänger:innen noch nicht ausdifferenziert ist (Krauser *et al.* 2018; Schiller 2017).

Der Untersuchungsablauf orientierte sich an der unter 3.2.2. geschilderten Reihenfolge und war bei allen Teilnehmenden derselbe. Die Studierenden wurden gebeten, gesund zum Termin zu erscheinen und andernfalls einen Ersatztermin zu vereinbaren. Für die wissenschaftliche

Analyse wurde die Einwilligung jeder/jedes Studienteilnehmer:in eingeholt. Die perzeptive Beurteilung des Stimmklangs der Studierenden wurde von zwei Sprecherziehern und einer Sprecherzieherin im Gruppenurteil im Nachgang zu den Screeningsitzungen gefällt. Als Kontrollhörer fungierte ein klinischer Sprechwissenschaftler. Für die Analyse wurde das Statistikprogramm IBM®SPSS® Version 25 benutzt.

4.2. Stichprobe

Die folgende Darstellung basiert auf einer Stichprobe der ersten 204 Studierenden, die das Stimmscreening-Angebot an der Universität Regensburg ab dem Wintersemester 2010/11 in Anspruch genommen haben. Die Rekrutierung erfolgte durch Bewerbung des Angebots „Stimmscreening für Lehramtsstudierende“ in Einführungsveranstaltungen zur „Didaktik der deutschen Sprache und Literatur“. Die Studienteilnahme erfolgte durch Selbstselektion.

Die untersuchte Stichprobe umfasst 167 weibliche (81,9%) und 37 männliche Studierende (18,1%). Von den 204 Studierenden gaben 198 ihr Alter an: Die Altersspanne erstreckt sich von 18 bis 47 Jahren mit einem durchschnittlichen Alter von $M = 22,56$ ($SD = 3,28$; $Median = 22$; $Modus = 21$) Jahren. 178 Studierende nannten ihr Fachsemester. Die Fachsemesterzahl erstreckt sich vom 1. Semester bis zum 13. Semester mit einer durchschnittlichen Semesteranzahl von $M = 3,72$ ($SD = 2,64$; $Median = 3$; $Modus = 1$).

Für die Berechnung des „Dysphonia Severity Index“ stehen aufgrund eines beschädigten Stimmfeld Datensatzes die Daten von 203 Studierenden zur Verfügung. Aufgrund fehlender Werte bei der Beantwortung aller 13 symptomassoziierten Fragen des konstruierten Fragebogens in der Untersuchung von Gegner (2021) können für die Berechnung des „Dysphonie-Index nach Friedrich“ die Daten von 198 Studierenden herangezogen werden.

4.3. Ergebnisse der „Dysphonia Severity Index“-Berechnung

Im Folgenden werden die Ergebnisse zu den Werten, die für die Berechnung des „Dysphonia-Severity-Index“ benötigt werden, berichtet. Dies sind im Einzelnen: höchste erzielte Frequenz ($F0_{max}$) der Singstimme, niedrigste Intensität (SPL_{min}) der Singstimme, die maximale Tonhaldedauer (MPT) und der Jitter-Wert in Prozentangabe (Wuyts *et al.* 1997, 2000). Tabelle 5 zeigt die deskriptive Statistik der genannten Werte.

		<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
F_{0max} in Hz	Männer	392,00	1188,30	696,57	149,39
	Frauen	392,00	1365,00	961,46	202,16
	Gesamt	392,00	1365,00	913,18	218,79
SPL_{min} in dB (A)	Männer	43,00	58,00	50,41	3,58
	Frauen	43,00	60,00	49,61	3,34
	Gesamt	43,00	60,00	49,76	3,39
MPT in Sekunden	Männer	11,55	42,40	23,47	7,96
	Frauen	7,60	38,85	17,71	5,30
	Gesamt	7,60	42,40	18,76	6,26
Jitter in Prozent	Männer	0,05	1,79	0,21	0,30
	Frauen	0,05	2,89	0,32	0,47
	Gesamt	0,05	2,89	0,30	0,45
DSI-Wert	Männer	1,50	9,30	5,79	1,69
	Frauen	0,30	11,50	6,53	1,98
	Gesamt	0,30	11,50	6,39	1,95

Bem.: Männer ($n = 37$); Frauen ($n = 166$); Gesamt ($N = 203$); *Min*: Minimum; *Max*: Maximum; *M*: Mittelwert; *SD*: Standardabweichung

Tabelle 5

Deskriptive Werte zur Berechnung des „Dysphonia Severity Index“ ($N = 203$).

Hinsichtlich der höchsten erzielten Frequenz bei der Singstimme erreichten die weiblichen Studierenden im Durchschnitt höhere Werte ($M_{weibl.} = 961,44$ Hz; $SD = 202,16$ Hz; $Range = 392$ Hz – 1365 Hz) als die männlichen Studierenden ($M_{männl.} = 696,57$ Hz; $SD = 149,39$ Hz; $Range = 392$ Hz – 1188,30 Hz). In Bezug auf den minimalen Schalldruckpegel der Singstimme erreichten weibliche Studierende durchschnittlich nur wenig geringere Werte ($M_{weibl.} = 49,61$ dB (A); $SD = 3,34$ dB (A); $Range = 43$ dB (A) – 60 dB (A)) als ihre männlichen Kommilitonen ($M_{männl.} = 50,41$ dB (A); $SD = 3,58$ dB (A); $Range = 43$ dB (A) – 58 dB (A)). Bei der aerodynamischen Messung verfügten die männlichen Studierenden durchschnittlich über eine längere maximale Phonationszeit ($M_{männl.} = 23,74$ s; $SD = 7,96$ s; $Range = 11,55$ s – 42,40 s) als die weiblichen Studierenden ($M_{weibl.} = 17,71$ s; $SD = 5,30$ s; $Range = 7,60$ s – 38,85 s). Bei der Jitter-Messung fielen weibliche Studierende im Durchschnitt über einen höheren Jitter-Wert in Prozent ($M_{weibl.} = 0,32$; $SD = 0,47$; $Range = 0,05$ – 2,89) als die männlichen Studierenden ($M_{männl.} = 0,21$; $SD = 0,30$; $Range = 0,05$ – 1,79) auf.

Anhand der eben berichteten Werte wird nach der unter 3.2.2. genannten Formel der DSI als multiparametrisches Verfahren zur Beschreibung der Stimmfunktion berechnet. Im Durchschnitt zeigten die weiblichen Studierenden hier höhere Werte ($M_{weibl.} = 6,53$; $SD = 1,98$; $Range = 0,30$ – 11,50) als die männlichen Studierenden ($M_{männl.} = 5,79$; $SD = 1,69$; $Range = 1,50$ – 9,30). Die Grenzwerte des DSI erstrecken sich von -5 für sehr pathologische Stimmen bis hin zu +5 für normale Stimmen, wobei diese Werte

auch deutlich über- bzw. unterschritten werden können. Im Durchschnitt liegen die Werte der weiblichen und männlichen Studierenden in dieser Untersuchung über dem Wert von +5. Allerdings zeigen die Minimalwerte, dass die Kennwerte für normale Stimmen auch teilweise deutlich unterschritten wurden. Nach dem lingWAVES Voice Diagnostic Center (VDC) Benutzerhandbuch (Tabelle 1) können die ermittelten Werte Intervallen zugeordnet und in folgende diagnostische Skala überführt werden:

Skala	Männer (n = 37)		Frauen (n = 166)		Gesamt (N = 203)	
	absolute Häufigkeit	Prozent	absolute Häufigkeit	Prozent	absolute Häufigkeit	Prozent
5	31	83,8%	144	86,7%	175	86,2%
4	4	10,8%	13	7,8%	17	8,4%
3	1	2,7%	5	3,0%	6	3,0%
2	1	2,7%	4	2,4%	5	2,5%
1	0	0,0%	0	0,0%	0	0,0%
0	0	0,0%	0	0,0%	0	0,0%

Bem.: 5 = keine Dysphonie (> 4,4); 4 = sporadische Dysphonie (+3,4 bis +4,3); 3 = leichte Dysphonie (+2,3 bis +3,3); 2 = zahlreiche Dysphonieperioden (+0,4 bis +2,2); 1 = sporadische, periodische Phonation (-1,9 bis +0,3); 0 = anhaltende Dysphonie (< -2,0).

Tabelle 6

Prozentuale Verteilung der untersuchten Stichprobe (N = 203) in Bezug auf die DSI-Klassen nach dem lingWAVES Voice Diagnostic Center (VDC) Benutzerhandbuch (Version 3.2 – 2017-09, S. 54) aus Gegner (2021, S. 297).

Demnach verfügten 86,2% der 203 Studierenden (Männer: 83,8%; Frauen: 86,7%) zum Untersuchungszeitpunkt über eine normale Stimmfunktion (keine Dysphonie). Durch eine sporadische Dysphonie für kurze Momente fielen 8,4% der Studierenden auf (Männer: 10,8%; Frauen 7,8%). Durch eine leichte Dysphonie 3,0% der Studierenden (Männer: 2,7%; Frauen: 3,0%). Zahlreiche Dysphonieperioden wurden bei 2,5% der Studierenden (Männer: 2,7%; Frauen: 2,4%) ermittelt. Niemand fiel zum Untersuchungszeitpunkt durch sporadische, periodische Phonation oder eine anhaltende Dysphonie auf.

4.4. Ergebnisse der adaptierten Dysphonie-Index-Berechnung nach Friedrich

Für eine bessere Lesbarkeit werden die einzelnen diagnostischen Bereiche für die Indexberechnung nach Friedrich (1998) separat berichtet (Tabelle 4): (1) auditive Stimmklangbeurteilung nach der RBH-Klassifikation; (2) Stimmumfang der Singstimme in HT; (3) Stimmodynamik zwischen leiser Sprechstimme und Rufen in dB; (4) maximale Tonhaltedauer in Sekunden; (5)

Subjektive Einschätzung anhand des konstruierten Fragebogens zu 13 symptomassozierten Beschwerden. Der Index entspricht dem arithmetischen Mittel.

4.4.1. Ergebnisse zur auditiven Stimmklangbeurteilung

Im Folgenden wird zuerst das Gruppenurteil zur Heiserkeitsbewertung der untersuchten Stichprobe durch die Expert:innengruppe aus Sprechwissenschaft und Sprecherziehung wiedergegeben.

Bereich	Punkte	Männer (<i>n</i> = 37)	Frauen (<i>n</i> = 161)	Gesamt (<i>N</i> = 198)
		Prozent	Prozent	Prozent
Auditive	0	73,0%	33,5%	40,9%
Stimmklang- beurteilung nach RBH	1	27,0%	54,7%	49,5%
	2	0,0%	11,8%	9,6%
	3	0,0%	0,0%	0,0%

Bem.: 0 = unauffällig, 1 = gering auffällig, 2 = mittelgradig auffällig, 3 = hochgradig auffällig.

Tabelle 7

Heiserkeitsbeurteilung der untersuchten Stichprobe (*N* = 198) aus Gegner (2021, S. 281); Gegner *et al.* (2024, S. 150).

Wie aus Tabelle 7 hervorgeht, wurde von den Expert:innen der Stimmklang bei 40,9% der 198 untersuchten Studierenden (Männer: 73,0%; Frauen: 33,5%) als euphon zum Untersuchungszeitpunkt beschrieben. Der Stimmklang von 49,5% (Männer: 27,0%; Frauen: 54,7%) wurde als geringgradig und von 9,6% (ausschließlich Frauen: 11,8%) als mittelgradig heiser eingestuft. Niemand fiel durch eine hochgradige Heiserkeit auf.

4.4.2. Ergebnisse zum Stimmumfang

In Tabelle 8 finden sich die erreichten Halbtöne bei der Untersuchung der Singstimme. Die Halbtonanzahl wird vom tiefsten zum höchsten möglichen Ton auf Basis der Singstimmleistung durch lingWAVES errechnet (ohne Unterscheidung zwischen lauter und leiser Singstimme).

Bereich	Punkte	Männer (<i>n</i> = 37)	Frauen (<i>n</i> = 161)	Gesamt (<i>N</i> = 198)
		Prozent	Prozent	Prozent
Stimmumfang	0	100%	95,0%	96,0%
der Singstimme	1	0,0%	5,0%	4,0%
in Halbtönen	2	0,0%	0,0%	0,0%
(HT)	3	0,0%	0,0%	0,0%

Bem.: Stimmumfang: 0 = > 24 HT, 1 = 24 – 18 HT, 2 = 17 – 12 HT, 3 = < 12 HT.

Tabelle 8

Erreichte Halbtonanzahl der Singstimme der untersuchten Stichprobe (*N* = 198) aus Gegner (2021, S. 281); Gegner *et al.* (2024, S. 150).

Von den 198 analysierten Studierenden erreichten 96,0% (Männer: 100%; *Range*: 26 HT – 47 HT; Frauen: 95,0%; *Range*: 25 HT – 41 HT) die geforderte Halbtonanzahl der Singstimme von > 24 HT. Lediglich 4,0% unterschritten diesen Wert (ausschließlich Frauen; *Range*: 19 HT – 24 HT). Niemand fiel durch einen geringeren Umfang der Singstimme als 19 HT auf.

4.4.3. Ergebnisse zur Stimmdynamik

Die Stimmdynamik wird nicht durch die Software ermittelt, sondern muss durch die Differenz des Schalldruckpegels der Rufstimme und des leisest möglichen Tons der Sprechstimme ermittelt werden. Die errechneten Werte wurden den Intervallen nach Friedrich (1998) zugeordnet (Tabelle 9).

Bereich	Punkte	Männer (<i>n</i> = 37)	Frauen (<i>n</i> = 161)	Gesamt (<i>N</i> = 198)
		Prozent	Prozent	Prozent
Stimmdynamik	0	83,8%	81,4%	81,8%
(dB) zwischen	1	16,2%	18,6%	18,2%
leiser Sprech-	2	0,0%	0,0%	0,0%
und Rufstimme	3	0,0%	0,0%	0,0%

Bem.: Stimmdynamik: 0 = > 45 dB, 1 = 45 – 35 dB, 2 = 34 – 25 dB, 3 = < 25 dB.

Tabelle 9

Erreichte Stimmdynamik zwischen leiser Sprech- und Rufstimme der untersuchten Stichprobe (*N* = 198) aus Gegner (2021, S. 281); Gegner *et al.* (2024, S. 150).

81,8% der untersuchten Studierenden (Männer: 83,8%; *Range*: 46 dB – 61 dB; Frauen 81,4%; *Range*: 46 dB – 64 dB) wiesen eine unauffällige Stimmdynamik von > 45 dB zwischen der leisen Sprech- und Rufstimme auf. 18,2% (Männer 16,2%; *Range*: 42 dB – 44 dB; Frauen 18,6%; *Range*: 38 dB – 45 dB) unterschritten diesen Wert. Niemand verfügte über eine niedrigere Stimmdynamik als 38 dB.

4.4.4. Ergebnisse zur Tonhaldedauer

Über die maximale Tonhaldedauer der Studierenden wurde bereits bei der DSI-Berechnung berichtet. Hier erfolgt die Klassierung der Ergebnisse in die Intervallskala nach Friedrich (1998):

Bereich	Punkte	Männer (<i>n</i> = 37)	Frauen (<i>n</i> = 161)	Gesamt (<i>N</i> = 198)
		Prozent	Prozent	Prozent
Tonhaldedauer	0	91,9%	64,6%	69,7%
	1	8,1%	28,0%	24,2%
	2	0,0%	7,5%	6,1%
	3	0,0%	0,0%	0,0%

Bem.: Tonhaldedauer: 0 = > 15 s, 1 = 15 – 11 s, 2 = 10 – 7 s, 3 = < 7s.

Tabelle 10

Erreichte Tonhaldedauer der untersuchten Stichprobe (*N* = 198) aus Gegner (2021, S. 281); Gegner *et al.* (2024, S. 150).

Von den 198 Studierenden erreichten 69,7% eine Tonhaldedauer von > 15 Sekunden (Männer: 91,9%; *Range*: 15,25 s – 42,40 s; Frauen: 64,6%; *Range*: 15,35 s – 38,85 s). 24,2% der Studierenden fielen durch eine verminderte Tonhaldedauer von 15 bis 11 Sekunden auf (Männer: 8,1%; *Range*: 14,30 s – 11,55 s; Frauen: 28,0%; *Range*: 14,95 s – 11,15 s). 6,1% der Studierenden (ausschließlich Frauen; *Range*: 10,95 s – 7,60 s) fielen durch eine deutlich verkürzte Tonhaldedauer von < 11 s auf.

4.4.5. Ergebnisse zur Selbsteinschätzung

Die wahrgenommene kommunikative Stimmbeeinträchtigung wurde durch die durchschnittliche Häufigkeitsangabe zu 13 symptomassozierten Beschwerden ersetzt. Tabelle 11 zeigt die durchschnittliche Beschwerdebhäufigkeit in der Einteilung nach Friedrich (1998).

Bereich	Punkte	Männer (n = 37)	Frauen (n = 161)	Gesamt (N = 198)
		Prozent	Prozent	Prozent
Subjektive Beschwerden	0	45,9%	36,0%	37,9%
	1	54,1%	59,0%	58,1%
	2	0,0%	5,0%	4,0%
	3	0,0%	0,0%	0,0%

Bem.: Subjektive Beschwerden (Durchschnitt aus 13 subjektiven, symptomassoziierten Beschwerden): 0 = nie, 1 = selten, 2 = manchmal, 3 = häufig;

Tabelle 11

Durchschnittliche Beschwerdehäufigkeit der untersuchten Stichprobe (N = 198) aus Gegner (2021, S. 281); Gegner *et al.* (2024, S. 150).

Demnach bemerken durchschnittlich 37,9% der Studierenden (Männer 45,9%; Frauen 36,0%) keine subjektiven Beschwerden. 58,1% (Männer 54,1%; Frauen 59,0%) klagten „selten“ und 4,0% der Studierenden „manchmal“ (ausschließlich Frauen) über symptomassoziierte Beschwerden. Niemand nahm durchschnittlich „häufig“ Beschwerden zum Untersuchungszeitpunkt war.

4.4.6. Ergebnisse zur Indexberechnung nach Friedrich

Der Index zur Bestimmung des Schweregrads einer Stimmstörung berechnet sich als arithmetisches Mittel aus den fünf Bereichen: auditive Stimmklangbeurteilung, Stimmumfang, Stimmdynamik, Tonhaltedauer und subjektive Beschwerden. Für die untersuchte Stichprobe ergibt sich folgendes Ergebnis (Tabelle 12):

Bereich	Punkte	Männer (n = 37)	Frauen (n = 161)	Gesamt (N = 198)
		Prozent	Prozent	Prozent
Dysphonie- Index (Summe/5)	0	94,6%	64,0%	69,7%
	1	5,4%	36,0%	30,3%
	2	0,0%	0,0%	0,0%
	3	0,0%	0,0%	0,0%

Bem.: Dysphonie-Index: 0 = keine Dysphonie, 1 = geringgradige Dysphonie, 2 = mittelgradige Dysphonie, 3 = hochgradige Dysphonie.

Tabelle 12

Dysphonie-Index Berechnung der untersuchten Stichprobe (N = 198) aus Gegner (2021, S. 281); Gegner *et al.* (2024, S. 150).

Laut der adaptierten Indexberechnung nach Friedrich (1998) wiesen zum Untersuchungszeitpunkt von den 198 in die Analyse einbezogenen

Studierenden 69,7% keine und 30,3% (Männer: 5,4%; Frauen: 36%) eine geringgradige Dysphonie auf. Eine mittel- oder hochgradige Dysphonie konnte bei niemandem festgestellt werden.

4.5. Zusammenfassung

In der Gegenüberstellung der Ergebnisse des „Dysphonia Severity Index“ und der adaptierten Indexberechnung des „Dysphonie-Index nach Friedrich“ (1998) fällt ein deutlicher Unterschied hinsichtlich der als stimm auffällig eingestuften Studierenden bei Gegner (2021) auf:

Während bei der DSI-Berechnung nach lingWAVES lediglich 13,9% der einbezogenen 203 Studierenden als stimmlich auffällig eingestuft werden (8,4%: sporadische Dysphonie; 3,0%: leichte Dysphonie; 2,5%: zahlreiche Dysphonieperioden), identifizierte die Indexberechnung nach Friedrich (1998) bei 30,3% der einbezogenen 198 Studierenden eine geringgradige Dysphonie. Bereits Nawka und Wirth (2008, S. 178) geben hinsichtlich des DSI zu bedenken, dass dieser „besonders bei funktionellen und psychogenen Dysphonien [...] durchaus im Normalbereich liegen kann, obwohl die Stimme nicht tragfähig und belastbar ist“. Diesen Nachteil scheint die Indexberechnung nach Friedrich durch die Integration einer Selbsteinschätzung und einer perceptiven Einschätzung des Stimmklangs durch die/den Untersuchenden auszuräumen. Gerade diese Kombination subjektiver und quasi-objektiver Untersuchungsmethoden erlaubt es, einen „zusammenfassenden, synoptischen Eindruck der stimmlichen Leistungsfähigkeit bzw. Beeinträchtigung“ (Friedrich 1998, S. 14) der untersuchten Person abzubilden. Obwohl die softwarebasierte Berechnung des DSI in der Untersuchungspraxis die praktikablere und ökonomische scheint, steht doch bei einer präventiven Untersuchung an einer Hochschule die Beurteilung der stimmlichen Leistungsfähigkeiten für einen zukünftigen Sprechberuf im Vordergrund. Aus diesem Grund scheinen rein instrumentell-akustisch erfasste Parameter wie beim DSI die stimmliche Leistungsfähigkeit nicht gänzlich abzubilden. Aus diesem Grund werden in der Stimmscreening-Sprechstunde an der Universität Regensburg auch die anderen in Tabelle 2 aufgeführten Untersuchungsparameter berücksichtigt, obwohl viele von ihnen nicht in ein multiparametrisches Verfahren einfließen. Bei der untersuchten Stichprobe in der Untersuchung von Gegner (2021) zeigte sich zudem, dass von den 203 Studierenden 33,5% (10 Männer; 58 Frauen) den Referenzwert von ≤ 50 dB (A) bei der leisen Sprechstimme (*Range*: 51 dB (A) – 58 dB (A)) überschritten. Hinsichtlich der Lage der indifferenten/ungespannten Sprechstimmlage fielen 25,1% der Studierenden auf (8 Männer, 43 Frauen). Sie überschritten die normgerechte Lage von 3 – 8 HT oberhalb der unteren Grenze des Stimmumfangs (*Range*: 9 HT – 15 HT). Als interessanter Nebenbefund kann angeführt werden, dass hinsichtlich der Vortragstimmlage (der

vorherrschenden Arbeitsstimmelage von Lehrkräften) 37,4% (16 Männer; 60 Frauen) der Studierenden durch eine zu starke Tonhöhenzunahme im Vergleich zur erreichten Lautstärke auffielen. Viele Studierende wurden lediglich höher, aber kaum oder gar nicht lauter. Dies kann auf einen unphysiologischen Gebrauch der lauten Sprechstimme hinweisen, der eine Überbelastung der Stimmfunktion im späteren Berufsalltag begünstigen kann (Gegner 2021).

Aus diesem Grund wäre es wünschenswert, dass weitere Instrumente zur Beurteilung der stimmlichen Leistungsfähigkeit entwickelt werden, um bereits im Studium stimmlich auffällige Studierende zu identifizieren und ihnen die Chance zu geben, sich bestmöglich auf die stimmlichen Anforderungen ihres späteren Berufes vorzubereiten.

5. Fazit und Ausblick

Gerade für Studierende ohne Kenntnis eines physiologischen Stimmgebrauchs (Meuret 2017) und mit mangelnder stimmlicher Leistungsfähigkeit erhöht sich das Risiko, im Laufe ihres späteren Berufslebens eine Stimmstörung zu entwickeln (Schneider *et al.* 2004). Eine Prognose, die besonders angesichts des gegenwärtigen Lehrkräftemangels Anlass zur Sorge gibt und zum Umdenken in der Lehrkräfteaus- und -weiterbildung anregt. So unterstreicht auch eine Umfrage unter bayerischen Lehrkräften im Schuldienst ($N = 499$), dass ca. 40% ihre Stimme an einem typischen Unterrichtstag als beeinträchtigt (leicht-, mittel- und hochgradig) wahrnehmen und sich eine bessere und kontinuierliche Ausbildung in den Bereichen Stimme und Sprechen wünschen (Fröhlich-Necker 2021). In vielen Teilen Deutschlands (Eidenmüller 2016) und in Italien (Angelillo *et al.* 2009) findet derzeit keine Vorbereitung von Lehramtsstudierenden auf ihren stimm- und sprechintensiven Beruf statt. Aus diesem Grund sollten Präventionsangeboten wie Stimm screenings und Stimmtrainingsprogramme bereits in die Curricula der Lehramtsstudiengänge integriert und auch für die weiteren Phasen der Lehrkräftebildung angedacht werden. Ergebnisse zu den positiven Auswirkungen von Stimmtrainingsprogrammen auf die Stimmqualität von Lehrkräften liegen bereits vor (Richter *et al.* 2017), wobei sich auch anhaltende Effekte zeigen (Nusseck *et al.* 2021). Herausfordernd bei der Implementierung von derartigen Trainingsprogrammen scheinen besonders der Mangel an Ressourcen und die Umsetzung eines flächendeckenden Angebots zu sein. Vor dem Hintergrund der Digitalisierung (Jude *et al.* 2020) an Schulen ist es nur folgerichtig, dieses Potential auch im Sinne der Lehrkräftegesundheit zu nutzen. Ziel eines Forschungsprojekts an der Universität Regensburg in Kooperation mit dem Arbeitsmedizinischen Institut für Schulen in Bayern (AMIS-Bayern) ist es daher, eine App zur Stimmhygiene und -prophylaxe für bayerische Lehrkräfte

zu entwickeln und zu evaluieren: „ReSt-Regensburger Stimmtraining“. Das interdisziplinäre Vorhaben der Fächer Educational Data Science, Sprechwissenschaft/Sprecherziehung und Medizinische Psychologie soll die momentan auf Präsenz ausgelegten Trainingsprogramme auch digital verfügbar machen und somit eine Versorgungs- und Betreuungslücke schließen.

Bionote: Gegner, Christian, Dr., Sprecherzieher Univ./DGSS., Wissenschaftlicher Mitarbeiter am Lehrgebiet Mündliche Kommunikation und Sprecherziehung, Mitglied des Communication and Voice Centers for Teachers (CoVoC-T) an der Universität Regensburg, Arbeitsschwerpunkte: Stimme und Sprechen im Lehrberuf, Stimmscreening für Lehramtsstudierende, Entwicklung und Evaluation des Regensburger Stimmtrainings (ReSt), Rede- und Gesprächsrhetorik, Phonetik.

Author's address: christian.gegner@ur.de

Literatur

- Angelillo M., Di Maio G., Costa G., Angelillo N. and Barillari U. 2009, *Prevalence of occupational voice disorders in teachers*, in “Journal of Preventive Medicine and Hygiene” 50 [1], pp. 26-32.
- Beise M. 2023, *Lehrermangel: Das Lehren der anderen*. <https://www.sueddeutsche.de/politik/lehrermangel-deutschland-europa-frankreich-italien-pisa-finnland-schweiz-ungarn-1.5746408> (16.01.2024).
- Brockmann-Bausser M. and Bohlender J. 2014, *Praktische Stimmdiagnostik. Theoretischer und praktischer Leitfaden* (= Forum Logopädie), Thieme, Stuttgart.
- Byeon H. 2019, *The Risk Factors Related to Voice Disorder in Teachers: A Systematic Review and Meta-Analysis*, in “International Journal of Environmental Research and Public Health” 16 [19], 3675. <https://doi.org/10.3390/ijerph16193675>.
- Christmann B., Meißner B. and von Laguna K. 2024, *Stimm-screening an der RWTH Aachen. Die Entstehung des obligatorischen Modulbausteins*, in Schnebel S., Grassinger R., Visotschnig M.S., Wiedenhorn, T. and Janssen M. (Hrsg.), *Begleitung und Beratung – Konzepte zur Unterstützung professioneller Entwicklung im Lehramtsstudium*, Waxmann, Münster, S. 119-130.
- Dejonckere P., Bradley P., Clemente P., Cornut G., Crevier-Buchman L., Friedrich G., Van de Heyning P., Remacle M. and Woisard V. 2001, *A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatics of European Laryngological Society (ELS)*, in “European Archives of Oto-Rhino-Laryngology” 258 [2], pp. 77-82.
- Eichel H.W. 2015, *HNO-Heilkunde, Phoniatrie und Pädaudiologie für Sprachtherapeuten*, Urban & Fischer, München.
- Eidenmüller C. 2016, *Zur Prävalenz von Stimmstörungen in Lehrberufen*, in „Praxis Sprache“ 61, S. 105-110.
- Friedrich G. 1998, *Stimm-diagnostik in der Praxis*, in „Logopädie/Diplom Logopädinnen, Verband für Wien, Niederösterreich und Burgenland“ 4, S. 8-16.
- Friedrich G. 2013, *Stimm-diagnostik*, in Friedrich G., Bigenzahn W. and Zorowka P. (Hrsg.), *Phoniatrie und Pädaudiologie. Einführung in die medizinischen, psychologischen und linguistischen Grundlagen von Stimme, Sprache und Gehör*, Huber, Bern, S. 67-84.
- Fröhlich-Necker J. 2021, *Selbsteinschätzung der Stimm-situation bayerischer Lehrer:innen* (= unveröffentlichte Masterarbeit an der Universität Regensburg).
- Gassull C., Casanova C., Botey Q. and Amador M. 2010, *The impact of the reactivity to stress in teachers with voice problems*, in “Folia Phoniatica et Logopaedica” 62 [1-2], pp. 35-39.
- Gegner C. 2020, *Stimm-screening zur Feststellung der stimmlichen Leistungsfähigkeit bei Lehramtsstudierenden an Universitäten. Ein Vorschlag zu möglichen Bestandteilen und Klassierung von Untersuchungsergebnissen*, in Kranich W. (Hrsg.), *Sprechwissenschaft heute* (= Sprache und Sprechen, Band 52), Schneider, Baltmannsweiler, S. 149-159.
- Gegner C. 2021, *Mündliche Kompetenzen von Lehramtsstudierenden: Stimmliche Leistungsfähigkeit, rederhetorische und sprechkünstlerische Kompetenzen bei künftigen Deutschlehrkräften*, Frank & Timme, Berlin.
- Gegner C. and Schilcher A. 2022, *Mündliche Performanz im Lehramtsstudium: eine Übungslücke? Zur Notwendigkeit stimm- und sprechpraktischer Anteile im Lehramtsstudium*, in Heins J., Kleinschmidt-Schinke K., Wieser D. and Wiesner E.

- (Hrsg.), *Üben. Theoretische und empirische Perspektiven in der Deutschdidaktik* (= SLLD, Band 5), S. 151-181. <https://doi.org/10.46586/SLLD.248>.
- Gegner C., Sabath C. and Schilcher A. (2024), *Stimm-screening für Lehramtsstudierende im Projekt KOLEG2 an der Universität Regensburg. Ein Angebot des Communication and Voice Centers for Teachers (CoVoC-T)*, in Schnebel S., Grassinger R., Visotschnig M.S., Wiedenhorn, T. and Janssen M. (Hrsg.), *Begleitung und Beratung – Konzepte zur Unterstützung professioneller Entwicklung im Lehramtsstudium*, Waxmann, Münster, S. 133-156.
- Hacki T. 1999, *Tonhöhen- und Intensitätsbefunde bei Stimmgeübten. Vergleichende Sprechstimmfeld-, Rufstimmfeld- und Singstimmfeldmessung*, in: „HNO“ 47, S. 809-815.
- Hahn V. 2009, *Myofunktionelle Störungen*, in Grohnfeldt M. (Hrsg.), *Lehrbuch der Sprachheilpädagogik und Logopädie, Band 2, Erscheinungsformen und Störungsbilder*, Kohlhammer, Stuttgart, S. 327-339.
- Hazlett D.E., Duffy O.M. and Moorhead S.A. 2011, *Review of the Impact of Voice Training on the Vocal Quality of Professional Voice Users: Implications for Vocal Health and Recommendations for Further Research*, in “Journal of Voice” 25 [2], pp. 181-191.
- Hillegeist K. and Thomas F. 2024, *Stimm- und Sprechscreening als Steuerungsinstrument für Lehramtsstudierende. Professionalisierung für den Sprechberuf*, in Schnebel S., Grassinger R., Visotschnig M.S., Wiedenhorn, T. and Janssen M. (Hrsg.), *Begleitung und Beratung – Konzepte zur Unterstützung professioneller Entwicklung im Lehramtsstudium*, Waxmann, Münster, S. 97-118.
- Hukelmann V. 2009, *Jede Stimme zählt. Eine Untersuchung zu stimmpräventiven Maßnahmen für Lehramtsstudierende an deutschen Hochschulen*, in „sprechen“ 29 [55], S. 31-44.
- Hwa Chen, S., Chiang S.-C., Chung Y.-M., Hsiao L.-C. and Hsiao T.-Y. 2010, *Risk Factors and Effects of Voice Problems for Teachers*, in “Journal of Voice” 24 [2], pp.183-192.
- Jude N., Ziehm J., Goldhammer F., Drachsler H. and Hasselhorn M. 2020, *Digitalisierung an Schulen – eine Bestandsaufnahme*, DIPF| Leibniz-Institut für Bildungsforschung und Bildungsinformation, Frankfurt am Main.
- Krauser E., Berschin G. and Seibert N. 2018, *Stimme und stimmliche Belastung im Selbstkonzept von Lehrkräften und Lehramtsstudierenden*, in Angerer P., Muth T. (Hrsg.), *DGAUM – 58. Wissenschaftliche Jahrestagung 2018. Arbeiten im Alter. Arbeit im Gesundheitswesen. Arbeit mit chronischen Erkrankungen*, S. 115-119. https://www.dgaum.de/fileadmin/pdf/Jahrestagung/2010-2018/DGAUM_2018_Kongressdokumentation.pdf (16.01.2024).
- Lange J. and Appel J. 2014, *Stimmig unterrichten. Eine Analyse der Sprecherziehung im Lehramtsstudium aus logopädischer Sicht*, in „sprechen“ 58, S. 45-58.
- Martins R.H., Pereira E.R., Hidalgo C.B. and Tavares E.L. 2014, *Voice disorders in teachers. A review*, in “Journal of Voice” 28 [6], pp. 716-724. <https://doi.org/10.1016/j.jvoice.2014.02.008>.
- Meuret S. 2017, *Welche Faktoren beeinflussen die Gesundheit der Pädagogenstimme?*, in Fuchs M. (Hrsg.), *Die Stimme im pädagogischen Alltag* (= Kinder- und Jugendstimme, Band 11), Logos, Berlin, S. 49-54.
- Nawka T. and Anders L.C. 1996, *Die auditive Bewertung heiserer Stimmen nach dem RBH-System. Doppel-Audio-CD mit Stimmbeispielen*, Thieme, Stuttgart.
- Nawka T., Franke I. and Galkin E. 2006, *Objektive Messverfahren in der Stimm-diagnostik*, in „Forum Logopädie“ 4 [20], S. 14-21.
- Nawka T. and Gonnermann U. 2003, *Stimmstörungsindex (SSI)*, in Gross M. (Hrsg.), *Aktuelle phoniatisch-pädaudiologische Aspekte*, Median, Heidelberg, S. 375-379.

- Nusseck M., Immerz A., Spahn C., Echternach M. and Richter B. 2021, *Long-term effects of a voice training program for teachers on vocal and mental health*, in “Journal of Voice” 35 [3], pp. 438-446.
- Richter B. and Echternach M. 2007, *Die kranke Sprech- und Singstimme. Ursachen aufspüren und richtig behandeln*, in „HNO-Nachrichten“ 6, S. 18-22.
- Richter B. and Echternach M. 2010, *Stimm diagnostik und -therapie bei Angehörigen stimmintensiver Berufe*, in „HNO“ 58 [4], S. 389-396.
- Richter B., Spahn C., Echternach M., Immerz A. and Nusseck M. 2017, *Evaluation eines Seminarangebots zur stimmlichen und mentalen Gesundheit im Lehramtsreferendariat – eine empirische Studie*, in Fuchs M. (Hrsg.), *Die Stimme im pädagogischen Alltag* (= Kinder- und Jugendstimme, Band 11), Logos, Berlin, S. 65-81.
- Rittich E. 2018, *Grundlagen der Gesundheitsförderung und Prävention*, in Rittich E., Tormin S. and Bock B. (Hrsg.), *Prävention von Stimmstörungen* (= Forum Logopädie), Thieme, Stuttgart, S. 39-68.
- Seidner W. and Eysholdt U. 2015, *Diagnostik*, in Wendler J., Seidner W. and Eysholdt U. (Hrsg.), *Lehrbuch der Phoniatrie und Pädaudiologie*, Thieme, Stuttgart, S. 105-138.
- Murgia S., Mekus T. and Bottalico P. 2022, *Intelligibility of dysphonic speech in primary schools*, in “Journal of the Acoustical Society of America” 151 [4]: A169. <https://doi.org/10.1121/10.0011002>.
- Schiller I. 2017, *Stimmstörungen bei Lehrkräften im Vorbereitungsdienst*, in Hannken-Illjes K., Franz K., Gauß E.-M., Könitz F. and Marx S. (Hrsg.), *Stimme, Medien, Sprechkunst* (= Sprache und Sprechen, Band 49), Schneider, Baltmannsweiler, S. 269-277.
- Schneider B., Cecon M., Hanke G., Wehner S. and Bigenzahn W. 2004, *Bedeutung der Stimmkonstitution für die Entstehung von Berufsdysphonien*, in „HNO“ 52 [5], S. 461-467.
- Schneider-Stickler B. and Bigenzahn W. 2013, *Stimm diagnostik. Ein Leitfaden für die Praxis*, Springer, Wien.
- Schultz-Coulon H.-J. 1990, *Stimmfeldmessung*, Springer, Heidelberg.
- Sick S 2019, *Stimmstörungen bei Lehrkräften und Lehramtsstudierenden. Teil 1: Zum Forschungshintergrund*, in „sprechen“ 36 [67], S. 67-94.
- Vertanen-Greis H, Löytyniemi E. and Uitti J. 2020, *Voice Disorders are Associated With Stress Among Teachers: A Cross-Sectional Study in Finland*, in “Journal of Voice” 34 [3]: P488.e1–488.e8. <https://doi.org/10.1016/j.jvoice.2018.08.021>.
- Voigt-Zimmermann S. 2017, *Auswirkungen der heiseren Stimme von Pädagogen auf die Leistungen von Kindern*, in Fuchs M. (Hrsg.), *Die Stimme im pädagogischen Alltag* (= Kinder- und Jugendstimme, Band 11), Logos, Berlin, S. 37-48.
- Wuyts F., Van de Heyning P., De Bodt M., Heylen L. And Remacle M. 1997, *Dysphonia severity index: a multiparameter approach for the assessment of voice quality*, in “LARYNX-1997”, pp. 73-76.
- Wuyts F., De Bodt M., Molenberghs G., Remacle M., Heylen L., Millet B., Van Lierde K., Raes J. and Van de Heyning P. 2000, *The Dysphonia Severity Index: An Objective Measure of Vocal Quality Based on a Multiparameter Approach*, in “Journal of Speech Language and Hearing Research” 43 [3], pp. 796-809.

LOS GLOSARIOS FINANCIEROS ELECTRÓNICOS EN ESPAÑOL Y PORTUGUÉS COMO INSTRUMENTOS DE EDUCACIÓN FINANCIERA EN LATINOAMÉRICA

CARLOS ROCHA OCHOA
UNIVERSIDAD DE ECONOMÍA DE VIENA

Abstract – This study presents a novel comparative analysis of lexicographic tools created within the framework of financial education and inclusion initiatives in Latin America. In recent years, central banks and financial authorities around the world have taken on the task of facilitating access to financial education, aligning with global initiatives like the OECD's International Network on Financial Education (INFE). Among other instruments aimed at enhancing citizens' financial literacy, monolingual glossaries of economic and financial terms can be found on the Internet portals of many Latin American central banks. The glossaries included in the study were selected based on their alleged use of simplified language, purpose, format, accessibility and subject, excluding those targeted at economic specialists. Moreover, to ensure comparability, the corpus focuses on glossaries with a minimum of ninety-nine defined terms, resulting in a set of five glossaries from the central banks of Argentina, Bolivia, Brazil, Mexico, and Paraguay. Similar glossaries from other Latin American central banks were excluded given that they had economic experts as their target public. The lexicographic resources were analyzed in light of a functional approach (Tarp 2000; Bergenholtz, Tarp 2002; Fuertes-Olivera, Tarp 2014), which places great attention not only on the structure and nature of the lemmas, but also on the author's goals and explanations, the accessibility, the methodology and the use of internet technologies. The findings reveal significant disparities in terms of the composition, accessibility and relevance of the electronic financial glossaries produced by central banks in Latin America and serve as a foundation for future advancements in the fields of electronic lexicography and terminology for promoting financial education.

Keywords: lexicography; terminology; financial education; financial glossaries.

1. Introducción

En las últimas dos décadas, la educación financiera, entendida aquí como el proceso de alfabetización financiera general, ha cobrado una relevancia cada vez mayor. Esto se debe, en parte, a las crisis financieras que han sacudido al sistema económico mundial y al desarrollo de los ideales de justicia social que proponen que la inclusión y la equidad social deben formar parte integral

de los planes de sostenibilidad económica.¹ Como respuesta a estos desafíos, y en concordancia con iniciativas globales como la llamada Red Internacional de Educación Financiera (OECD-INFE, por sus siglas en inglés), desarrollada por la Organización para la Cooperación y el Desarrollo Económicos (OCDE) en 2008, los gobiernos de distintos países comenzaron a establecer planes y estrategias afines a través de sus bancos centrales y autoridades financieras.²

En este contexto, con el objetivo de llegar a un público amplio, se han creado no solo cursos y materiales informativos como guías y folletos, sino también herramientas digitales. En las plataformas en línea y las aplicaciones móviles se encuentran ahora juegos interactivos, videos educativos, artículos informativos e incluso simuladores financieros que buscan ayudar al público a aprender sobre cuestiones financieras. Ahora bien, si se buscara definir una tipología de dichos recursos en línea para la educación financiera, resultaría idóneo recurrir a la clasificación ofrecida por Mas (2018, p. 121):

1. recursos con información general y consejos,
2. calculadoras y herramientas de planificación,
3. sitios con datos y listados de acceso abierto al público,
4. vídeos explicativos, y
5. cursos estructurados de educación financiera.

En la primera categoría, de índole más general, destaca la presencia de los sitios web de las autoridades financieras y los organismos de protección del consumidor. Es en estos sitios web donde se encuentran los glosarios que constituyen el objeto de estudio de este trabajo de investigación. En estos sitios web se tiene acceso libre y gratuito a una serie de glosarios económico-financieros, cuyo objetivo cardinal es brindar definiciones y explicaciones simplificadas de conceptos relacionados con la economía y las finanzas. Este tipo de glosarios, que suele presentarse como parte de las herramientas de educación financiera y acceso a la información, ha sido poco investigado, no solo en castellano sino también en otros idiomas,³ por lo que resulta oportuno realizar un análisis de su funcionalidad, estructura y contenido desde un punto de vista lexicográfico y terminológico.

En estas páginas se presenta un análisis comparativo de cinco glosarios financieros electrónicos del ámbito latinoamericano (cuatro en español y uno

¹ Véase, por ejemplo, la publicación *Social Sustainability in Development: Meeting the Challenges of the 21st Century* (2023), del Banco Mundial. Disponible en <https://openknowledge.worldbank.org/entities/publication/0e9063cb-c2d7-401f-9f13-da64a24fc773> (último acceso: 20-05-2023).

² España fue uno de los países pioneros al lanzar en 2008 una iniciativa conjunta de educación financiera a cargo del Banco Central de España y la Comisión Nacional del Mercado de Valores.

³ El estudio más parecido es el de Preite (2015), en francés, sobre los glosarios electrónicos de la crisis económica, todos ellos creados por periódicos franceses y disponibles en sus sitios web.

en portugués) que han sido compilados por bancos centrales con objetivos y temáticas afines. Por esta razón se han excluido aquellos glosarios dirigidos a especialistas en temas (macro)económicos, dando como resultado un corpus de cinco glosarios que suman casi dos mil entradas y provienen de los bancos centrales de Argentina, Bolivia, Brasil, México y Paraguay. El análisis se basa en una perspectiva funcional, tomando en cuenta no solo la estructura sino también elementos metalexigráficos como objetivos y explicaciones proporcionadas por los autores, así como la accesibilidad o ruta de acceso, la metodología y el grado de uso de las tecnologías de Internet. A pesar de tener objetivos y públicos similares, los resultados revelan una significativa variación entre los distintos glosarios en términos de estructura y de contenido, así como en cuanto a su pertinencia en función del objetivo y público meta. Finalmente, a partir de todo ello, se ofrecen pautas a seguir para la elaboración de recursos lexicográficos de este tipo en el futuro.

2. Educación financiera y recursos lexicográficos

2.1. Iniciativas de educación financiera

Es posible distinguir una serie de hitos que han marcado el desarrollo reciente de las iniciativas de educación financiera a nivel global. El primero de ellos se registró en 2002 con el proyecto integral lanzado por la OCDE, pero que solo alcanzó la madurez con el establecimiento de la *International Network on Financial Education* (INFE), también a cargo de la OCDE, en 2008. Para respaldar esta iniciativa, tanto la misma OCDE⁴ como el Banco Mundial⁵ han venido publicando directrices para el diseño e implementación de Estrategias Nacionales de Inclusión Financiera (ENIF), como se les llama en Argentina, Ecuador y Paraguay, o Estrategias Nacionales de Educación Financiera (ENEF), como se les llama en Brasil y Bolivia.⁶ Actualmente, la INFE cuenta con la participación de organismos gubernamentales de 131 países,

⁴ Véase, por ejemplo, el manual *National Strategies for Financial Education: OECD/INFE Policy Handbook* (2015), disponible en <https://www.oecd.org/finance/National-Strategies-Financial-Education-Policy-Handbook.pdf> (último acceso: 15-05-2023).

⁵ Véase, por ejemplo, la nota técnica *Building a Financial Education Approach: A Starting Point for Financial Sector Authorities* (2021), disponible en <https://documents1.worldbank.org/curated/en/883771629960132032/pdf/Building-a-Financial-Education-Approach-A-Starting-Point-for-Financial-Sector-Authorities-Financial-Inclusion-Support-Framework-Technical-Note.pdf> (último acceso: 15-05-2023).

⁶ En México se le ha llamado Política Nacional de Inclusión Financiera (PNIF).

incluyendo tanto bancos centrales como ministerios de finanzas y de educación.⁷

En América Latina, al igual que en otras regiones del mundo, se han creado programas de educación financiera, muchos de los cuales han sido formalizados en las llamadas Estrategias Nacionales de Educación Financiera (ENEF).⁸ Sin embargo, diferentes encuestas han revelado que, en general, los niveles de educación financiera en la región son bajos en una gran parte de la población (García *et al.* 2013; Mejía, Rodríguez 2016), evidenciando así una necesidad urgente de mejorar la alfabetización financiera en la región. Si bien es verdad que los programas y el grado de desarrollo varían ampliamente, los bancos centrales figuran como los principales impulsores de estos programas en los países latinoamericanos, seguidos por el sector financiero privado (García *et al.* 2013; Roa *et al.* 2014). Por otro lado, un aspecto característico de las iniciativas de la región es la llamada inclusión financiera como objetivo de la educación financiera, como lo demuestra el Programa de Inclusión Financiera (2018) del Banco de Desarrollo de América Latina. Lo anterior resulta natural si se toma en cuenta que la región presenta niveles bastante bajos de inclusión en el sistema financiero formal en comparación con países industrializados, esto debido a una documentación insuficiente y a la falta de confianza en las instituciones financieras.⁹

En este sentido, se ha insistido en que la inclusión financiera es una vía hacia el desarrollo en la región latinoamericana, pues está relacionada con la reducción de brechas económicas, la disminución de la pobreza, una mayor productividad y bienestar económico (OCDE y CAF 2020). Por desgracia, la creación de programas de inclusión o educación financiera en América Latina ha sido más tardía que en otras regiones del mundo, con la notable excepción de Brasil. El país lusófono puso en marcha la llamada *Estratégia Nacional de Educação Financeira* por decreto nacional en 2010, solo dos años después de la creación de la INFE de la OCDE y el Plan de Educación Financiera de España (2008). E incluso más meritorio es quizá el hecho de que la estrategia brasileña haya sido lanzada un año antes que el *Plano Nacional de Formação Financeira* de Portugal, sentando de esta manera un precedente en el ámbito lusófono.

⁷ Cifras actuales disponibles en el sitio web de OCED/INFE <https://www.oecd.org/financial/education/oecd-international-network-on-financial-education.htm> (último acceso: 15-05-2023).

⁸ Según el Global Findex 2021, publicado por el Banco Mundial, el 41,67% de los adultos en América Latina y el Caribe aún carece de acceso a una cuenta bancaria. Disponible en <https://www.worldbank.org/en/publication/globalfindex> (último acceso: 15-05-2023).

⁹ Según el Global Findex 2021, publicado por el Banco Mundial, el 41,67% de los adultos en América Latina y el Caribe aún carece de acceso a una cuenta bancaria. Disponible en <https://www.worldbank.org/en/publication/globalfindex> (último acceso: 15-05-2023).

En el extremo contrario del espectro se sitúan países como las vecinas Bolivia y Argentina o Guatemala, quienes han postergado hasta el final o el cambio de década el lanzamiento de una estrategia nacional de inclusión y educación financiera semejante. En Bolivia, país que se unió de manera más tardía, el banco central presentó un documento de trabajo para discutir una posible estrategia nacional de inclusión financiera en diciembre de 2016,¹⁰ pero esta sólo se concretizó en 2022. A fin de ofrecer una panorámica más amplia del desarrollo histórico de la cuestión, la Tabla 1 presenta el año de puesta en marcha de las primeras estrategias de inclusión financiera (IF) y educación financiera (EF) en doce países latinoamericanos seleccionados:¹¹

País	IF	EF
Argentina	2019	2019
Brasil	2012	2010
Bolivia	2022	2022
Chile	N/A	2017
Colombia	2016	2017
Ecuador	2012	2012
Guatemala	2019	2019
Honduras	2015	2015
México	2016	2017
Paraguay	2014	2015
Perú	2015	2017
Uruguay	2014	2012

Tabla 1

Primeras estrategias de inclusión y educación financiera en América Latina.

2.2. Conceptos y definiciones

Desde un punto de vista terminológico, llama la atención que, en la literatura disponible de las organizaciones internacionales mencionadas anteriormente, el término “educación financiera” (*financial education*) suele ir acompañado por el término “conocimiento financiero” (*financial literacy*), o bien por sus sinónimos “cultura financiera” (*financial culture*) y “competencia financiera” (*financial capability*). El segundo término parece ser de uso mucho menos frecuente en castellano debido quizá a la vaga traducción del término *literacy* como “alfabetización”. Esto explicaría por qué, cuando se menciona, suele llevar el mismo significado que “educación”. En todo caso, la educación o

¹⁰ Disponible en <https://www.bcb.gob.bo/webdocs/publicacionesbcb/2018/07/58/29%20-%20Hacia%20una%20estrategia%20nacional%20de%20inclusi%C3%B3n%20financiera%20en%20Bolivia.pdf> (último acceso: 15-05-2023).

¹¹ Datos de OCDE y CAF 2020: 9, complementada con datos propios para Bolivia.

alfabetización financiera suele definirse de manera similar a otros tipos de educación, es decir, como un desarrollo de habilidades:

Proceso por el cual los consumidores/inversionistas mejoran su comprensión de los productos financieros, los conceptos y los riesgos y desarrollan las habilidades y confianza para ser más conscientes de los riesgos y oportunidades financieras, tomar decisiones informadas, saber a dónde ir para obtener ayuda y ejercer cualquier acción eficaz para mejorar su bienestar económico. (OCDE 2005; traducción de Roa y Carvallo 2018, p. 77)

En esta definición puede advertirse el papel de la comprensión no solamente de productos sino también de conceptos financieros con el fin de desarrollar las habilidades necesarias para alcanzar o mantener el bienestar económico. Sin embargo, los conocimientos y las habilidades son solo uno de los varios componentes que la educación financiera contempla, ya que se trata más bien de una mezcla de toma de conciencia, conocimientos, habilidades, actitudes y comportamientos (OCDE 2011, p. 3). Pero si nos centramos en el conocimiento, como sucede en otros campos más o menos especializados del saber humano, la adquisición de elementos terminológicos es un paso ineludible como parte del proceso de comprensión de conceptos y desarrollo de habilidades.

En el área en cuestión suele hablarse de conceptos básicos de economía y finanzas. No obstante, existe una línea continua y difusa entre conceptos económicos y conceptos financieros, puesto que las dos áreas se entrelazan y se complementan. En lo fundamental, los conceptos económicos habrán de referirse a fenómenos y principios relativos a la producción, la distribución y el consumo de bienes y servicios en una economía. Mientras que, por el otro lado, los conceptos financieros se centran en las actividades y herramientas relacionadas con la gestión del dinero, como la inversión, el ahorro, el crédito y los mercados financieros. A pesar de todo esto, resulta imposible negar que muchos conceptos financieros tienen una base económica subyacente, ya que están influenciados por factores macroeconómicos y políticas económicas.

Es justamente debido a esta superposición entre conceptos económicos y financieros que se ha planteado la necesidad de una comprensión integral para tener una visión global y sólida de la realidad financiera y económica.¹² Por lo tanto, los conceptos relativos a la gestión de las finanzas personales, incluyendo lo relacionado a salarios, compras, consumo de servicios, ahorro, créditos, inversiones y seguros,¹³ irán de la mano de conceptos económicos

¹² Véanse, por ejemplo, Kaminski y Friebe (2012), Davies (2015) y Fuhrmann *et al.* (2015).

¹³ Estas son las áreas comprendidas en el *Test of Financial Literacy* (TFL) del American Council of Economic Education, disponible en <https://www.councilforeconed.org/wp-content/uploads/2016/10/TFL-manual-final-040416.pdf> (último acceso: 15-05-2023).

como el producto interior bruto (PIB), la inflación, el presupuesto, la política fiscal y el tipo de cambio. Desde un punto de vista epistemológico, se podría decir entonces que la educación financiera en general es un componente de la educación económica, es decir, un todo integrado a nivel macro que implica el entendimiento de la sociedad más allá del nivel del individuo (Kaminski, Friebel 2012). De manera similar lo ha explicado también el Banco Central del Paraguay en su glosario, aunque con una sintaxis un tanto peculiar:

Educación económica y financiera: Proceso mediante el cual los individuos mejoran su comprensión de los conceptos y productos financieros con el fin de desarrollar capacidades, valores y habilidades necesarios para tomar decisiones financieras responsables, y comprendan [*sic*] los efectos que los cambios en los principales indicadores macroeconómicos pueden generar en su propio bienestar. (BCP, *Glosario Financiero*, 2018)

Por otra parte, más allá de su semejanza con lo planteado anteriormente, esta definición nos sirve para ejemplificar una característica lingüística recurrente en las definiciones contenidas en los glosarios de los bancos centrales: frases muy largas que dificultan la comprensión por parte del lector. El fenómeno se discute con mayor detalle en el apartado 3.3., donde se incluyen algunos otros ejemplos e implicaciones prácticas.

2.3. Los glosarios financieros electrónicos

Si, como hemos dicho, la terminología y la comprensión de conceptos son uno de los pilares de la educación financiera, los glosarios financieros, al ser repositorios de terminología, han de concebirse con un propósito y un público claro a fin de poder desempeñar un papel relevante. Algunos bancos centrales y otras instituciones financieras los presentan claramente en el marco general de sus planes de educación financiera, es decir, como obras de referencia para ayudar a los ciudadanos, entendidos como usuarios de servicios financieros, a comprender y utilizar correctamente el léxico de la economía y las finanzas personales. Este es el caso de los glosarios financieros de los bancos centrales de Argentina, Bolivia (en coautoría con la ASFI: Autoridad de Supervisión del Sistema Financiero), Brasil, México y Paraguay. En los demás casos, se trata de glosarios menos básicos destinados a facilitar la comprensión de las publicaciones del banco central en cuestión (por ejemplo, los de Chile y Perú) que presentan estadísticas macroeconómicas y datos técnicos sobre el estado de las finanzas nacionales.

Por otro lado, los cinco glosarios que constituyen el corpus del presente estudio tienen en común una preferencia por definiciones claras y precisas, al menos en palabras de sus autores, e incluyen una variedad de temas que han de permitir a las personas familiarizarse con conceptos clave como tasas de interés, inflación, activos financieros, préstamos e inversiones, entre otros. La

estructura de los glosarios es alfabética y el número de entradas oscila entre 99 (Banco Central del Paraguay) y 845 (Banco Central de Bolivia). Dos de ellos están disponibles como parte del sitio web del banco central (Argentina y México) mientras que los otros deben ser descargados en formato PDF. En la Tabla 2 se ofrece un listado con los cinco glosarios financieros, destacando la institución responsable, el título, año de publicación, número de entradas y formato de acceso:

País	Institución	Título	Año	Entradas	Formato
ARG	BCRA	<i>Diccionario financiero</i>	2017	529	Web
BOL	BCB/ASFI	<i>Glosario de términos económico financieros</i>	2019	845	PDF
BRA	BCB	<i>Glossário Simplificado de Termos Financeiros</i>	2013	114	PDF
MEX	BANXICO	<i>Glosario</i>	2017	349	Web
PRY	BCP	<i>Glosario financiero</i>	2018	99	PDF

Tabla 2
Glosarios financieros de bancos centrales en Latinoamérica.

En principio, un glosario consiste en un listado de términos utilizados en un campo específico de conocimiento como la medicina, la informática, o las finanzas. Además, a diferencia de una obra lexicográfica de carácter general, su función es proporcionar explicaciones y definiciones concisas de términos específicos y suele ser más limitado en cuanto a la cantidad de términos. Si el glosario o diccionario está disponible en formato digital, suele denominarse “diccionario electrónico” o “glosario electrónico”. Sin embargo, los glosarios o diccionarios electrónicos no siempre están disponibles en formatos digitales que faciliten la consulta y acceso a la información como los sitios web o las aplicaciones móviles. Además, en contra de lo que podría pensarse, no siempre incluyen funciones de búsqueda, enlaces o referencias cruzadas, ejemplos de uso u otras características interactivas que puedan facilitar la consulta y el aprendizaje de los términos incluidos.

En efecto, a través de los glosarios escogidos para este estudio puede comprobarse que algunos glosarios electrónicos son esencialmente versiones digitalizadas de los glosarios en papel, ya que están disponibles en el formato de documento digital PDF. Sánchez y Cantos (2011, p. 253) mencionaban ya en la segunda década del presente siglo que el cambio del diccionario en papel al diccionario electrónico no ha traído consigo cambios sustanciales. A pesar de esto, admitían los mismos autores, las versiones electrónicas de los glosarios o diccionarios brindan ventajas en términos de su portabilidad, accesibilidad, velocidad de recuperación, coste y precio, espacio, potencial de circulación y uso responsable de recursos. Otros autores, sin embargo, han concluido que un recurso lexicográfico de tales características, es decir, que

no hace un uso productivo de las tecnologías informáticas disponibles en la actualidad, no debe siquiera llamarse diccionario o glosario electrónico.¹⁴

3. Análisis lexicográfico y metalexicográfico

3.1. Objetivos y explicaciones metalexicográficas

A manera de recapitulación de lo expresado en el apartado anterior, los glosarios incluidos en este estudio han sido seleccionados bajo los siguientes criterios: son monolingües y del ámbito financiero o económico-financiero, se presentan como parte de iniciativas nacionales de educación financiera, han sido creados por o en colaboración con un banco central en América Latina y están disponibles de forma electrónica y gratuita.

La Teoría Funcional de la Lexicografía, propuesta por Bergenholtz y Tarp (2002),¹⁵ sirve como base teórica al centrarse en los aspectos prácticos y funcionales de los recursos lexicográficos orientados al usuario. Según esta teoría, las personas que consultan un glosario o un diccionario no requieren información general sino un tipo específico de información que depende tanto del tipo de usuario como del tipo concreto de situación en la que se produce esa necesidad. En este sentido, se pueden identificar cuatro funciones o situaciones de uso: comunicativas, cognitivas, operativas e interpretativas (Tarp 2008). Siguiendo este marco conceptual, los glosarios financieros monolingües aquí analizados son, como se ha explicado, parte de la educación financiera y buscan cubrir una necesidad cognitiva. En otras palabras, la consulta está destinada a ayudar al usuario a aprender sobre un tema o disciplina: la económico-financiera.

La misma teoría sirve para abordar aspectos metalexicográficos por medio de una serie de criterios. En este sentido, Fuertes Olivera y Tarp (2014) recomiendan examinar la visión del autor, las funciones, la ruta de acceso, las tecnologías de Internet utilizadas, la metodología lexicográfica, la información sobre actualizaciones, la participación de expertos y la selección de datos, además de los costos de producción y los costos de información. A continuación, se proporcionan los datos relevantes recogidos para los glosarios seleccionados.

Argentina: *Diccionario financiero*

El glosario del Banco Central de la República Argentina está disponible en www.bcra.gob.ar > BCRAyVos > Usuarios Financieros > Información al

¹⁴ Véase, por ejemplo, Fuertes Olivera y Niño Amo (2011) o Tarp (2012).

¹⁵ Ver Tarp (2008) para una revisión histórica de los orígenes de la teoría.

usuario financiero > Diccionario Financiero,¹⁶ y se trata de un listado en tres columnas al que puede accederse por letra en orden alfabético a través de un enlace. Los autores lo presentan como un “completo diccionario de términos bancarios y financieros”, para ayudar a los usuarios a interpretar de manera correcta los textos de esta temática. Además, afirman que “las definiciones vertidas [...] se encuentran redactadas en un lenguaje simple y resumido”. No se explica si el glosario es o ha sido actualizado ni si ha habido un grupo de expertos, ya sea lexicográficos o económicos, detrás del mismo. Sin embargo, las entradas incluyen una columna donde aparece la fuente, como se muestra en la Figura 1, y la existencia de esta página está documentada desde junio de 2017 según el Internet Archive,¹⁷ de modo que se ha tomado este año como la fecha de publicación para los fines de este trabajo de investigación.

A|B|C|D|E|F|G|H|I|J|K|L|M|N|Ñ|O|P|Q|R|S|T|U|V|W|X|Y|Z|#|

TÉRMINO	DEFINICIÓN	FUENTE
ABA:	Asociación de Bancos de la Argentina.	SEDESA
ABAPRA:	Asociación de Bancos Públicos y Privados de la Argentina	SEDESA
ABE:	Asociación de Banca Especializada	SEDESA

Figura 1

Captura de pantalla del *Diccionario financiero* del BCRA.

Bolivia: *Glosario de términos económico financieros*

El glosario del Banco Central de Bolivia y la Autoridad de Supervisión del Sistema Financiero (ASFI) está disponible en la página principal del sitio de la ASFI, www.asfi.gob.bo,¹⁸ para descargarse gratuitamente en formato PDF. Como se observa en la Figura 2, se trata de un listado en dos columnas y, en el prefacio, el Ministro de Economía lo ubica dentro del proceso de inclusión y educación financiera. Además, explica que contiene términos de la “actual estructura del sistema financiero” y “términos utilizados en la evaluación de la economía nacional” y que han sido definidos “con palabras sencillas y claras” (BCB/ASFI 2019, p. 8). En la primera página se especifica que estamos ante la segunda edición de este glosario (noviembre de 2019), pero no se incluye la fecha de la primera edición. Por último, el presidente del

¹⁶ Disponible en https://www.bcr.gov.ar/BCRAyVos/Herramientas_Diccionario_Financiero.asp (último acceso: 30-05-2023).

¹⁷ Disponible en <https://web.archive.org> (último acceso: 30-05-2023).

¹⁸ Disponible en https://www.bcr.gov.ar/BCRAyVos/Herramientas_Diccionario_Financiero.asp (último acceso: 30-05-2023).

Banco Central de Bolivia explica que ha sido un trabajo conjunto del banco central, de la antes mencionada ASFI y del Ministerio de Economía y Finanzas Públicas.

A

Acción. Instrumento de inversión de participación en proporción a lo aportado en la sociedad anónima (capital) que representa la parte alícuota del capital de una sociedad, concediendo a su titular derechos y obligaciones según correspondan a los accionistas, en directa proporción al monto de su inversión.

Actividad económica del deudor. Es la principal actividad que genera la fuente de repago del crédito.

Actividad financiera ilegal o no autorizada. Actividad efectuada por personas naturales o jurídicas, que sin previa autorización de ASFI realizan con carácter masivo y habitual actos propios de las entidades de intermediación financiera y/o de servicios financieros complementarios, así como la difusión de publicidad de actividades reservadas para las entidades financieras.

Figura 2

Captura de pantalla del *Glosario de términos económico financieros* del BCB/ASFI.

Brasil: *Glossário Simplificado de Termos Financeiros*

El glosario del Banco Central de Brasil está disponible en www.bcb.gov.br > Acesso à informação > Cidadania Financeira > Biblioteca (vídeos e documentos),¹⁹ para descargarse en formato PDF. En el documento se ofrece una lista numerada en una columna, como se observa en la Figura 3, y en el prefacio se insiste también en el importante papel de la terminología dentro de la educación financiera.²⁰ De manera bastante similar a los glosarios antes citados, se explica que contiene “definições de palavras, siglas e expressões relacionadas a produtos e serviços do mercado financeiro, utilizando linguagem cotidiana” y que es el resultado de un “esforço de simplificação” (BCB 2013, p. 5). Se menciona que estamos ante la primera edición (noviembre de 2013), que parece ser la única disponible hasta la fecha de redacción de este artículo. Por último, se ofrece una lista con los nombres de los miembros del equipo técnico que fue responsable por la elaboración del glosario.

¹⁹ Disponible en https://www.bcb.gov.br/content/cidadaniafinanceira/documentos_cidadania/Informacoes_gerais/glossario_cidadania_financeira.pdf (último acceso: 30-05-2023).

²⁰ “A educação financeira voltada para a tomada de decisões autônomas envolve a compreensão da linguagem do mundo financeiro.” (Banco Central do Brasil 2013, p. 5).

1. A prazo

Pagamento em data posterior à compra. Desse modo, o comprador não paga no momento da compra, mas fica com uma dívida (o dever de pagar no futuro), e o vendedor, com um crédito (o direito de receber no futuro).

Exemplo de uso: Essa loja aceita pagamento a prazo?

Ver também: à vista, dívida.

2. À vista

Pagamento no ato, imediatamente.

Exemplo de uso: Se eu comprar a televisão com pagamento à vista, a loja me dará um desconto.

Ver também: a prazo.

Figura 3

Captura de pantalla del *Glossário Simplificado de Termos Financeiros* del BCB.

México: *Glosario*

El glosario del Banco de México está disponible en el subportal de educación financiera Educa, www.educa.banxico.org.mx > Recursos > Glosario,²¹ y se trata de un listado en una columna al que puede accederse por letra en orden alfabético a través de un enlace, como se muestra en la Figura 4. No se ofrece explicación alguna por parte de los autores, pero la página principal del portal Educa indica que ofrece “información y material educativo para fortalecer tus competencias económicas y financieras”. No se indica si el glosario es o ha sido actualizado ni si ha habido un grupo de expertos, ya sea lexicográficos o económicos, detrás del mismo. Sin embargo, la existencia de esta página está documentada desde agosto de 2017 según el Internet Archive, de manera que se ha tomado este año como la fecha de publicación para los fines de esta investigación.

²¹ Disponible en http://educa.banxico.org.mx/recursos_banxico_educa/glosario.html (último acceso: 30-05-2023).

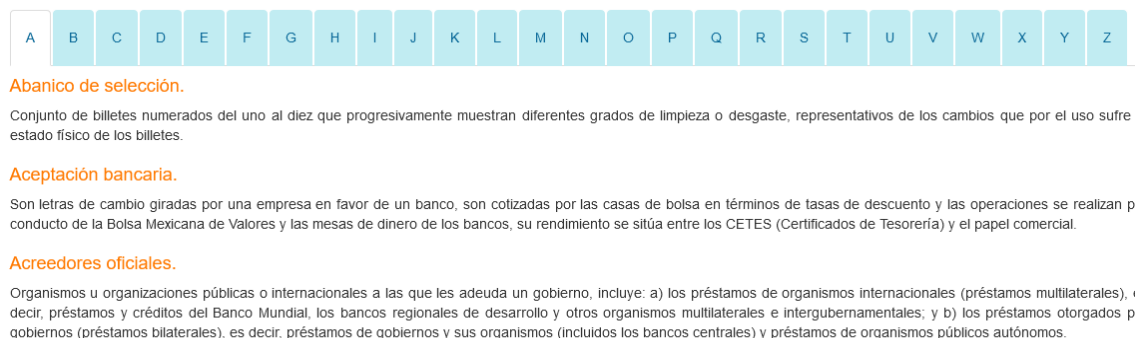


Figura 4
Captura de pantalla del *Glosario* del Banco de México.

Paraguay: *Glosario financiero*

El glosario del Banco Central del Paraguay está disponible en la sección para la inclusión y la educación financieras de su sitio web, www.bcp.gov.py > Inclusión Financiera > BCP Educa > Lee los materiales,²² para descargarse en formato PDF. El documento incluye una lista de “colaboradores expertos en educación financiera” en la ficha técnica y, como se aprecia en la Figura 5, el glosario se presenta como un listado enumerado en dos columnas. En general, es patente el esfuerzo por vincularlo con la Estrategia Nacional de Educación Financiera, pero también el énfasis en explicar el lenguaje como “elemento de inclusión y cohesión social, [así] como también de exclusión para quienes no alcanzan tal comprensión” (BCP 2018, p. 5). En el prefacio se dice también que la metodología de elaboración involucró un proceso de comparación de los términos financieros en materiales de educación e inclusión financiera de otros países, como Perú, Colombia, México, Brasil, España, Estados Unidos y Reino Unido, además de materiales de la OCDE y la ASBA (Asociación de Superintendencias Bancarias de América Latina) (Ibid.). Aún más inusual es la explicación sobre la fase de investigación, aunque no se explica en realidad si se ha contado con la participación de expertos lexicográficos:

Los responsables de la ENIF y de la Red de Microfinanzas del Paraguay se reunieron para determinar los conceptos claves que serían investigados y puestos a consideración, además de definir la participación de expertos que luego colaborarían en las etapas de discusión y validación de los términos finales. (Ibid., p. 6)

²² Disponible en https://www.bcr.gov.ar/BCRAyVos/Herramientas_Diccionario_Financiero.asp (último acceso: 30-05-2023).

- 1. ACREEDOR.** Es la persona física o jurídica que otorga un crédito o préstamo.
- 2. AHORRO.** Parte del ingreso que se guarda para utilizar en el futuro, es una decisión consciente sobre el uso de los ingresos. Todas las personas, independientemente del monto de sus ingresos, pueden ahorrar.
- 3. AHORRO FORMAL.** Es cuando el ahorro se realiza en una institución financiera (Bancos, Financieras o Cooperativas)
- 4. AHORRO INFORMAL.** Es cuando el ahorro se realiza a través de cualquier mecanismo fuera de las instituciones financieras.
- cuentas y/o acceder a otros servicios tales como consultas de saldo, transferencias de fondos y otros definidos por la entidad autorizante.
- 6. BANCA COMUNAL:** Es una organización de personas que bajo la figura del apoyo mutuo realizan operaciones de créditos individuales con una institución o entidad financiera y son garantizadas por el grupo, generalmente son personas no incluidas en el sistema financiero formal.
- 7. BANCA MÓVIL.** Es un tipo de servicio de banca web que ofrecen las instituciones financieras y que permite a los usuarios realizar ciertas operaciones financieras desde cualquier dispositivo móvil.

Figura 5

Captura de pantalla del *Glosario financiero* del BCP.

3.2. Composición, definiciones y elementos lexicográficos

En esta sección se procede a analizar la composición léxica de los glosarios seleccionados, que en conjunto suman un total de 1936 entradas. Para ello se distingue entre seis distintas categorías de unidades léxicas: unidades léxicas simples (ULS) (aquellas formadas por un solo segmento o palabra), unidades léxicas compuestas, complejas o poliléxicas (ULC) (formadas por más de una palabra), siglas, abreviaturas, símbolos y anglicismos, en caso de haberlos. Un rasgo general, y hasta cierto punto predecible, es la presencia dominante de unidades léxicas compuestas en relación con las unidades léxicas simples. Al observar la Tabla 3 es posible comprobar que más de dos terceras partes o, dicho de otro modo, siete de cada diez entradas (69,58%) en la totalidad del corpus son unidades léxicas compuestas. Este porcentaje sería aún mayor si se excluyeran las siglas, que en el caso argentino suponen un 43,3% del total.

	ARG	BOL	BRA	MEX	PRY	Total	%
Entradas	529	845	114	349	99	1936	100
ULS	70	111	52	68	32	333	17,20
ULC	214	733	56	280	64	1347	69,58
Siglas	229	1	6	1	3	240	12,40
Abreviaturas	14	0	0	0	0	14	0,72
Símbolos	2	0	0	0	0	2	0,10
Anglicismos	15	7	1	6	6	35	1,81

Tabla 3

Composición léxica de los glosarios financieros seleccionados.

En realidad, el *Diccionario financiero* del BCRA no solo destaca por la abundancia de siglas, sino también por las definiciones para las mismas. En la gran mayoría de los casos se trata de asociaciones y organismos nacionales o internacionales para los que la definición es únicamente el nombre, es decir, el sintagma nominal al cual se refieren las siglas.²³ Asimismo, destaca como el único glosario en el que se han incluido abreviaturas (EE.UU., mill., p.p., RRHH, entre otras) e incluso un par de símbolos (€, \$R). En el resto de los glosarios, por el contrario, la presencia de siglas es mínima y la aparición de abreviaturas y símbolos es nula. Por último, el examen de las entradas ha permitido identificar que al menos tres de ellas aparecen repetidas, como lo ilustra el ejemplo de la Figura 6 a continuación:

IPC GBA:	Índice de Precios al Consumidor del Gran Buenos Aires	BCRA
IPC GBA:	Índice de Precios al Consumidor del Gran Buenos Aires	BCRA
IPC Interior:	Índice de Precios al Consumidor del Interior del país	BCRA
IPC Nacional:	Índice de Precios al Consumidor Nacional	BCRA

Figura 6

Ejemplo de entradas repetidas en el *Diccionario financiero* del BCRA.

Por otra parte, la proporción de anglicismos es bastante baja en la totalidad del corpus y en el glosario brasileño es casi nula, pues solo figura *leasing*.²⁴ Este último glosario destaca también por tener una proporción de unidades léxicas compuestas (49,1%) muy por debajo del promedio (69,58%) y ofrece unidades simples y unidades compuestas casi a partes iguales.

Ahora bien, al examinar los elementos lexicográficos de los diferentes glosarios, resulta evidente que solo en el glosario del Banco Central de Brasil se ha contado con algún tipo de asesoramiento de especialistas lexicógrafos. En el *Glossário Simplificado de Termos Financeiros* se proporcionan no solo términos y definiciones, sino también sinónimos, ejemplos de uso y términos relacionados, es decir, referencias cruzadas, bajo la indicación *Ver também*. Aunque también tenemos referencias cruzadas en los glosarios de los bancos centrales de Argentina y Paraguay, no hay ni sinónimos ni ejemplos de uso ni parece haber un trabajo lexicográfico considerable detrás de los mismos. La Tabla 4 presenta una vista rápida de los elementos lexicográficos presentes o ausentes en los distintos glosarios, donde también se comprueba que solo dos

²³ La más conocida de las siglas incluidas para los lectores argentinos deberá ser, sin duda, CABA: Ciudad Autónoma de Buenas Aires.

²⁴ Este anglicismo figura también en los glosarios del BCRA (Argentina) y del BCB/ASFI (Bolivia).

de los glosarios se han centrado en el léxico financiero (Fi), mientras que los demás incluyen una mezcla de éste y del léxico macroeconómico (Ec):

País	Tema	Entr.	Form.	Sinón.	Ref.C.	Ej.Uso	Búsqu.
ARG	Ec/Fi	529	Web	No	Sí	No	No
BOL	Ec/Fi	845	PDF	No	No	No	N/A
BRA	Fi	114	PDF	Sí	Sí	Sí	N/A
MEX	Ec/Fi	349	Web	No	No	No	No
PRY	Fi	99	PDF	No	Sí	No	N/A

Tabla 4

Elementos lexicográficos en los glosarios financieros seleccionados.

Un análisis más minucioso sobre los campos temáticos revela que el glosario del Banco Central del Paraguay, el más breve de todos, es un caso único al incluir términos actuales relativos a la tecnología financiera (p. ej. *big data*, *bitcoin*, *blockchain*, *fintech*). Además, a pesar de no contar con los elementos lexicográficos del glosario brasileño, es el único que ofrece una ontología que puede facilitar un aprendizaje estructurado, por ejemplo, un breve desglose de las “partes de un contrato de seguro” (asegurado, beneficiario, siniestro, indemnización, riesgos no asegurables) y tipos de seguro (de automóviles, de vida, inclusivo, médico, social) debajo de la entrada “seguro”.

A nivel de las definiciones existe también un alto grado de variación entre los distintos glosarios, lo cual puede observarse de manera más clara al comparar las unidades léxicas que tienen en común. Al tomar como ejemplo el término “cuenta corriente”, disponible en todos los glosarios menos el del Banco Central de la República Argentina, se puede advertir que definición de mayor longitud, como la del glosario brasileño, es más clara debido al uso de frases cortas y lenguaje sencillo. Por el contrario, una definición más breve, como la del glosario boliviano, puede resultar confusa debido a la densidad de términos y el uso de frases subordinadas. La definición paraguaya, aunque similar a esta, muestra un esfuerzo por simplificar. Por último, la definición en el glosario mexicano deja claro que se está hablando de macroeconomía y no de finanzas personales, de manera que no es realmente comparable:

BOL: Contrato realizado entre una entidad financiera supervisada con una persona natural o jurídica, mediante el cual éste último realiza la apertura de una cuenta de depósito a la vista que utiliza generalmente el giro de cheques para el movimiento de los fondos. (BCB/ASFI 2019, p. 47)

BRA: É o nome pelo qual a conta de depósitos à vista é popularmente conhecida. É a maneira mais comum de manter dinheiro em uma instituição financeira. Funciona como um cofre, em que o cliente deposita seu dinheiro e pode ter acesso a serviços como pagamento de contas, saques, transferências, emissão de cheques e realização de compras com cartão de débito, entre outros. Para guardar o dinheiro e oferecer serviços como os citados, a

instituição financeira pode cobrar tarifas, mas o cliente é quem escolhe se prefere pagar uma tarifa individualizada por serviço que utilizar ou uma tarifa única que dá direito a um pacote de serviços. (BCB 2013, p. 17)

MEX: Componente de la balanza de pagos donde se registra el comercio de bienes y servicios y las transferencias unilaterales de un país con el exterior. Las principales transacciones de servicios son los viajes y el transporte; y los ingresos y pagos sobre inversiones extranjeras. Las transferencias unilaterales se refieren a regalías hechas por los particulares y el gobierno a los extranjeros, y a regalías recibidas de extranjeros. La exportación de bienes y servicios y el ingreso de transferencias unilaterales entran en la cuenta corriente como créditos (con signo positivo) porque llevan el recibo de pagos provenientes de extranjeros. Por otra parte, la importación de bienes y servicios y el otorgamiento de transferencias unilaterales se registran como débitos (con signo negativo) porque significa el pago a extranjeros.²⁵

PRY: Es un producto ofrecido exclusivamente por los bancos de acuerdo con las condiciones previamente pactadas, mediante el cual se deposita dinero en el mismo, pudiendo realizar transacciones en cualquier momento a través del uso de cheques y otros mecanismos. (BCP 2018, p. 10)

Sin embargo, las definiciones también pueden ser demasiado simplificadas o breves, como lo ejemplifica una comparación entre las definiciones para el término “acción” en el glosario argentino y en el glosario boliviano. Como se aprecia en la cita textual más abajo, la definición en el glosario boliviano no solo es más larga, sino también menos clara debido al léxico y a la sintaxis. Por el contrario, la definición en el glosario argentino es incluso más breve y menos informativa que la de un diccionario general como el DLE (Diccionario de la Lengua Española):

ARG: Valor negociable que representa una participación en una sociedad comercial.²⁶

BOL: Instrumento de inversión de participación en proporción a lo aportado en la sociedad anónima (capital) que representa la parte alícuota del capital de una sociedad, concediendo a su titular derechos y obligaciones según correspondan a los accionistas, en directa proporción al monto de su inversión. (BCB/ASFI 2019, p. 21)

DLE: Título valor que representa una parte alícuota en el capital de una sociedad mercantil y que da derecho a una parte proporcional en el reparto de

²⁵ Disponible en http://educa.banxico.org.mx/recursos_banxico_educa/glosario.html (último acceso: 30-05-2023).

²⁶ Disponible en https://www.bcra.gob.ar/BCRAyVos/Herramientas_Diccionario_Financiero.asp (último acceso: 30-05-2023).

beneficios y a la cuota patrimonial correspondiente en la disolución de la sociedad.²⁷

La comparación anterior también parece sugerir que la definición de la RAE pudo haber servido de inspiración para la definición en el glosario boliviano, aunque con una sintaxis un poco menos clara. La razón principal de ello es el empleo de nominalizaciones, gerundios, participios y oraciones subordinadas. Al final, los ejemplos presentados son solo un acercamiento a los patrones en la composición y las definiciones que exhiben los distintos glosarios.

4. Conclusiones

Las explicaciones proporcionadas por los bancos centrales indican implícita o explícitamente que los glosarios financieros promueven la igualdad de acceso a la información financiera al proporcionar una base común de terminología utilizada en el sector. Esta uniformidad en la terminología fomenta una mayor transparencia y evita confusiones o malentendidos en la interpretación, lo que a su vez promueve una toma de decisiones más informada y responsable por parte de los individuos en relación a su bienestar financiero.

Sin embargo, a pesar del auge reciente de las iniciativas de educación e inclusión financiera y la tendencia a incluir información terminológica como parte de las herramientas disponibles, no parece haber una visión uniforme ni consistencia en cuanto al contenido temático y el nivel de especialización de las definiciones. Solo los glosarios financieros del BCRA (Argentina) y del BCP (Paraguay) ofrecen alguna información sobre las fuentes y no se ofrece ni se encuentra evidencia de la contribución de expertos lexicógrafos, con la excepción del glosario brasileño y, en menor medida, del glosario paraguayo que parece haber sido inspirado por el primero.

Desde un punto de vista técnico, el uso de las tecnologías de Internet es bastante limitado, incluso en el caso de los glosarios de Internet (página web) como BCRA y BANXICO, pues ninguno cuenta con un campo de búsqueda a pesar de albergar varios cientos de entradas en cada caso. En términos de la composición, el primero de ellos es más un glosario de siglas y abreviaturas que un glosario financiero, y ambos ofrecen más léxico macroeconómico que de finanzas personales. A juzgar por la estructura y las definiciones incluidas, solo el glosario brasileño y el glosario paraguayo entregan lo que prometen y, por ello, pueden servir para el desarrollo de futuras herramientas en el campo de la educación financiera desde un punto de vista lexicográfico.

Este análisis constituye un primer esfuerzo que puede ser utilizado para futuros avances en el campo de la lexicografía especializada con el propósito

²⁷ Disponible en <https://dle.rae.es/acción> (último acceso: 30-05-2023).

de promover una educación financiera más efectiva en castellano o en otros idiomas. Sin embargo, también es necesario llevar a cabo estudios empíricos que ayuden a determinar el grado de comprensibilidad de las definiciones y la adaptación a las necesidades de los usuarios. Para ello, es esencial contar con la participación no solo de lexicógrafos sino también de expertos financieros e informáticos. Solo si se emplean las tecnologías disponibles en la actualidad que facilitan la consulta y el aprendizaje por parte del usuario podrá hablarse de “glosarios financieros electrónicos” en un sentido estricto.

Bionote: Carlos O. Rocha-Ochoa graduated in International Studies and German (University of Guadalajara), and holds Master’s degrees in Scandinavian Studies (University of Iceland) and Applied Linguistics (Korea University). He is currently a Research and Teaching Associate (pre-doc) in Romance Languages for Business Purposes (*Wirtschaftsromanistik*) at the Vienna University of Economics and Business, where he teaches Spanish Business Communication. His main research interests lie within the fields of terminology, lexicography, translation, pragmatics, and languages for specific purposes (LSP). His recent projects focus on terminological anglicisms in Spanish financial lexicography, forthcoming in *Terminàlia*, and pragmatic variation in digital business communication across the Spanish-speaking world, forthcoming in *Studia Linguistica Romanica*. He has previously published research on Scandinavian philology in Spanish, English, and Portuguese, and gathered over fifteen years of experience as a general, technical, and literary translator from English, German, and Scandinavian languages into Spanish.

Author’s address: carlos.rocha.ochoa@wu.ac.at

Bibliografía

- Barron P., Cord L., Cuesta J., Espinoza S., Larson G., Woolcock M. 2023, *Social Sustainability in Development: Meeting the Challenges of the 21st Century*. New Frontiers of Social Policy, World Bank, Washington DC.
- Bergenholtz H. und Tarp S. 2002, *Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen*, in „Lexicographica“ 18, pp. 253-263.
- Davies P. 2015, *Towards a framework for financial literacy in the context of democracy*, in “Journal of Curriculum Studies” 47 [2], pp. 300-316.
- Fuertes Olivera A. and Niño Amo M. 2011, *Internet Dictionaries for Communicative and Cognitive Functions: El Diccionario Inglés-Español de Contabilidad*, in Fuertes Olivera, P.A. and Bergenholtz, H. (eds.), *E-Lexicography: The Internet, Digital Initiatives and Lexicography*, Bloomsbury, London/New York, pp. 251-574.
- Fuertes Olivera P.A. and Tarp S. 2014, *Theory and Practice of Specialised Online Dictionaries*, De Gruyter, Berlin/New York.
- Fuhrmann B., Cechovsky N. und Riess J. 2021, *Finanzbildung – „Life Skill“ zur Erschließung der Welt?*, in „Pädagogische Rundschau“ 75 [1], pp. 47-58.
- García N., Grifoni A., López J.C. y Mejía D. 2013, *La educación financiera en América Latina y el Caribe. Situación actual y perspectivas*, en “Serie Políticas Públicas y Transformación Productiva” 12, CAF, Caracas.
- Kaminski H. und Friebel S. 2012, *Finanzielle Allgemeinbildung als Bestandteil der ökonomischen Bildung*, Institut für Ökonomische Bildung, Oldenburg.
- Leroyer P. 2017, *Dictionaries for text reception*, in P.A. Fuertes-Olivera (ed.), *The Routledge Handbook of Lexicography*, Routledge, New York, pp. 250-266.
- Mas I. 2018, *Using Broadband to Enhance Financial Education*, in D. Lee Kuo Chuen and R. Deng (eds.), *Handbook of Blockchain, Digital Finance, and Inclusion. Vol. 1: Cryptocurrency, FinTech, InsurTech, and Regulation*, Academic Press, London, pp. 91-130.
- Mejía D. y Rodríguez G. 2016, *Determinantes socioeconómicos de la educación financiera. Evidencia para Bolivia, Colombia, Ecuador y Perú*, en “Serie de Políticas Públicas y Transformación Productiva” 23, CAF, Bogotá.
- OCDE (Organización para la Cooperación y el Desarrollo Económicos) 2005, *Improving Financial Literacy: Analysis of Issues and Policies, Financial Market Trends*, vol. 2. OCDE, París.
- OCDE (Organización para la Cooperación y el Desarrollo Económicos) 2011, *Measuring Financial Literacy: Core Questionnaire in Measuring Financial Literacy: Questionnaire and Guidance Notes for conducting an Internationally Comparable Survey of Financial literacy*. OCDE, París.
- OCDE (Organización para la Cooperación y el Desarrollo Económicos) 2020, *Overview of the OECD International Network on Financial Education*, OCDE, París. <https://www.oecd.org/financial/education/oecd-infe-overview.pdf>
- OCDE (Organización para la Cooperación y el Desarrollo Económicos) y CAF (Banco de Desarrollo de América Latina) 2020, *Estrategias nacionales de inclusión y educación financiera en América Latina y el Caribe: retos de implementación*, OCDE, París.
- Preite C. 2015, *Dictionnaires Internet et vulgarisation des termes de la crise économique et financière*, dans Paissa P., Rigat F., Vittoz M.B. (éd.), « Dans l’amour des mots: chorale(s) pour Mariagrazia », Edizioni dell’Orso, Alessandria, pp. 449-464.

- Roa M.J., Alonso G., García N. and Rodríguez D. 2014, *Financial Education and Inclusion in Latin American and the Caribbean*. CEMLA and Banco de la República de Colombia, Mexico City and Bogotá.
- Roa M.J. y Carvallo O. 2018, *Inclusión financiera y el costo del uso de instrumentos financieros formales: las experiencias de América Latina y el Caribe*, Banco Interamericano de Desarrollo, Washington D.C.
- Sánchez A. and Cantos P. 2011, *e-Dictionaries in the Information Age: The Lexical Constellation Model (LCM) and the Definitional Construct*, in Fuertes-Olivera P.A. and Bergenholtz H. (eds.), *E-Lexicography: The Internet, Digital Initiatives and Lexicography*, Bloomsbury, London/New York, pp. 251-574.
- Tarp S. 2008, *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*, Niemeyer, Tübingen.
- Tarp S. 2012, *Theoretical challenges in the transition from lexicographical p-works to e-tools*, in Granger S. and Paquot M. (eds.), *Electronic Lexicography*, Oxford University Press, Oxford, pp. 107-118.

Glosarios

Banco Central de Bolivia (BCB) y Autoridad de Supervisión del Sistema Financiero (ASFI) 2019 (segunda edición), *Glosario de términos económico financieros*, BCB/ASFI, La Paz.

https://www.asfi.gob.bo/images/EDUCACION_FINANCIERA/DOCS/Glosario_de_T%C3%A9rminos_Econ%C3%B3mico_Financieros.pdf

Banco Central de la República Argentina (BCRA) 2017, *Diccionario financiero*. BCRA, Ciudad Autónoma de Buenos Aires.

https://www.bcra.gob.ar/bcrayvos/Herramientas_Diccionario_Financiero.asp

Banco Central del Paraguay (BCP) 2018, *Glosario financiero*. BCP, Asunción.

<https://enif.paraguay.gov.py/storage/app/uploads/public/5ae/0c5/f1b/5ae0c5f1b12ad216401736.pdf>

Banco Central do Brasil (BCB) 2013, *Glossário Simplificado de Termos Financeiros*, BCB, Brasília.

https://www.bcb.gov.br/content/cidadaniafinanceira/documentos_cidadania/Informacoes_gerais/glossario_cidadania_financeira.pdf

Banco de México (BANXICO) 2017, *Glosario*, BANXICO, Ciudad de México.

https://educa.banxico.org.mx/recursos_banxico_educa/glosario.html

DICCIOCHO

An 18th Century Lexicographic Portal

ELENA CARPI
UNIVERSITÀ DI PISA

Abstract – This article deals with the *Dicciocho* lexicographic portal, dedicated to eighteenth-century Spanish dictionaries: *Dicciocho* was born from the desire to make the word list of the second edition of *Autoridades* available to the scientific community. The manuscript of this dictionary, considered lost, was recently found in the archives of the Royal Spanish Academy (RAE), and it can now be consulted online on the RAE website. Later, the portal was expanded to include materials related to the first edition of *Autoridades* and the essential *Diccionario castellano con voces de Ciencias y Artes* by the Jesuit Father Terreros y Pando. To exemplify the use of the portal, an analysis of the artistic entries in the two *Autoridades* will be proposed.

Keywords: 18th c. lexicography; 18th c. lexicology; RAE; *Dicciocho*; *Diccionario de autoridades*.

The best way out is always through.
(Robert Frost “A Servant to Servants”,
1914).

1. Introduction

Spanish lexicography took its first steps in the centuries before the eighteenth, with the bilingual dictionaries of Alonso de Palencia and Nebrija¹. The importance of the latter's work is evident in relation to the following bilingual lexicographic dictionaries, such as those by Cristobal de las Casas (1570), by Richard Percyvall (1599), by Cesar Oudin (1616) and by Lorenzo Franciosini (1620). In 1611, Sebastián de Covarrubias published the *Tesoro de la lengua Castellana, o Española*, the first monolingual Spanish dictionary, which the academics of the RAE cite in the *Prólogo* as the “obra tan grande, que ha servido à la Academia de clara luz” (RAE 1726: I).

¹ Alonso de Palencia published in 1490 the *Universal Vocabulario en latín y romance*; Elio Antonio de Nebrija is the author of the *Lexicon hoc est Dictionarium ex sermone latino in hispaniensem* (1492) and of the *Dictionarium ex hispaniensi in latinum sermonem* (1495).

Despite the great importance of these precedents, the creation of the RAE in 1713 with the main task of working on the dictionary of the Spanish language marks the starting point of the Spanish lexicographical trajectory. In fact, for academics,

no era decente [...] que logrando la fortuna de encontrarla [la lengua española] en nuestros días tan perfecta, no eternizassemos en las prensas su memoria, formando un Diccionario al exemplo de las dos celebradísimas Academias de París y Francia. ([Cassani] 1726a: xi, § 4)

However, the eighteenth century is not only the century that gives Spain the *Diccionario grande* (“large”² dictionary), in imitation of the one that Italy and France already possessed³; it is also that in which the tradition of bilingual and trilingual dictionaries continues⁴, and the *Diccionario castellano con las voces de ciencias y artes* of Terreros is published. The latter relates to the lexicographical tradition of the Jesuits, which in France, from 1701 to 1771, experienced a moment of utmost importance with the so-called *Dictionnaire de Trevoux*⁵.

The lexicographical richness that, as shown, distinguishes the eighteenth century, is the main reason why the *Grupo de investigación Dicciónarística del siglo XVIII* decided to bring together in a portal its research on the lexicography of the Age of Enlightenment. *Dicciocho. Portal lexicográfico del XVIII*, the subject of this article, aims to provide a common space for study and reflection on this fundamental moment. In fact, it includes insights concerning the first edition of *Diccionario de autoridades* (1726-39), of the manuscript of the second *Diccionario de autoridades* (1780-1829)⁶ and of the *Diccionario castellano con las voces de ciencias y artes* (1786-88) by Esteban de Terreros y Pando. In the section “Other dictionaries”⁷, a list of mono- and multilingual repertoires presents the

² The academics of the RAE called *Diccionario grande* the dictionary containing quotations from well-known Spanish authors, as opposed to the *Diccionario común o vulgar* (common or vulgar dictionary) in one tome.

³ The *Vocabolario degli Accademici della Crusca* was printed in Venice in 1612 and the *Dictionnaire de l'Académie Française* was published in 1694. It is well known that the RAE, soon after its foundation, requested copies of their Statutes to the two sister academies. In the prologue of *Diccionario*, the academics themselves acknowledge that “para la formación de este Diccionario se han tenido presentes los de las Lénguas extrangéras, y especialmente el Vocabulario de la Crusca de Florencia” (RAE: 1726: II). Regarding the relations between the RAE and the *Crusca*, see Carpi and Carriscondo (2023).

⁴ <https://dicciocho.org/2020/10/29/otros-diccionarios/>

⁵ On the influence of the *Dictionnaire de Trevoux* on the *Diccionario* of Father Terreros, see Carpi (2021).

⁶ From now on called DA1 and DA2.

⁷ *Otros diccionarios*.

lexicographical panorama of the eighteenth century with hyperlinks for consultation.

Dicciocho is a repository in constant development, in which the materials are exhibited as a gallery, with thematic rooms depending on different criteria: chronological (archaisms, recently introduced words), technical (specialized entries, science and the arts), regional (dialectisms, lexicon of American Spanish) and marked lexicon (subject categories of lexicon) according to the registers, the frequency of use and the affected diastatic varieties. In this article, we will focus only on the part of the portal which concerns the dictionaries of the *Autoridades*, devoting special attention to the lexicon of the Arts.

2. Diccionario de autoridades (DA1) (1726-39)

This section of the portal includes the different thematic fields of DA1 (1726-39), the list of which we propose in Annex 1. As can be seen, there are more than one hundred thematic fields covering the most diverse areas related to Sciences, Arts and Crafts. This proves the interest of the academics of the RAE in specialized lexicon, despite what is stated in the Prologue:

De las voces propias pertenecientes à Artes liberales y mecánicas ha discurrido la Académia hacer un Diccionario separado, quando este se haya concluído: por cuya razón se ponen solo las que han parecido mas comunes y precisas al uso, y que se podían echar menos. (RAE 1726)

The dictionary is not organized with a standard marker system, a feature that implies a conscientious work of analysis when assigning a lexical item to a certain field. The study of the definitions showed that the technical words are defined by a great variety of syntagmas that function as diatechnical markers; for this reason, we started from the query of the *Diccionario de autoridades*⁸ with keywords such as *término*, *term.*, *voz*, *palabra*. This first step allowed us to enrich the keywords and compile the list of the various thematic fields, shown in Annex 1. With a view to the future rationalisation of the various subject areas, we have established fields that can be integrated into a higher one. As an example, we offer those that accompany the architecture entries, similar to those we have found in other areas. These statements, headed by *término*, *voz* or *term.*, in some cases refer to the concept:

término de arquitectura, term. de architectúra, voz de la arquitectura, es voz de arquitectura, es voz de la architectúra, en la architectúra es, es en la architectúra, tiene uso en la architectúra, en este sentido es voz mui común en

⁸ <https://apps2.rae.es/DA.html>

la arquitectura, en este sentido es mui usado en la arquitectura, es termino usado en la arquitectura, usase en la arquitectura, usase mucho en la arquitectura, se llama en la arquitectura, en la arquitectura llaman, dícese en la arquitectura de.

Otherwise, they refer to the extralinguistic world, as in the cases of *llaman tambien los arquitectos* or *entre los arquitectos se llama*.

The importance of having defined the thematic fields of DA1 lies in the fact that it leads to further reflection on this point starting from DA2, when the RAE includes a functional marking in the microstructure.

3. Diccionario de autoridades (DA2) (1737-1829)

Regarding the second manuscript edition of the *Diccionario de Autoridades*, a revised and corrected supplement to the entire work was announced in the fifth volume of the first edition (1737). This plan was later substituted by the project of a second edition of which only the first volume was published, although scholars worked on it until 1829.

Annex 2, in the tables 5 and 6, presents the chronological sequence followed for the elaboration of the new dictionary and the physical documentation of these materials, deposited in the Archive of the RAE and found in recent years: thirty-seven units of documents divided into nine folders plus twenty-eight notebooks with sheets, which contain a total of 13,943 sheets *in folio*.

During 2019, a part of the *Dicciocho* project team worked in the archive on recently discovered materials not yet published on the RAE portal; In *Dicciocho* we present the results of this research, which refers to the material prepared in the 18th century up to the letter F⁹.

One of the characteristics of this second edition is the use of abbreviations to mark, among other types of information, the technical nature of the words. Only the new additions, compared to DA1 (1726-39), are listed here, marked as pertaining to a certain field according to the formalized procedure followed for DA1. In this case, also, the thematic fields are miscellaneous: pottery, pharmacy, animal breeding, gardening, horse handling, mining, hairdressing, rhetoric, cooperage, winery, zoology¹⁰.

⁹ <https://dicciocho.org/2020/10/30/tesauro/> <24/10/2023>

¹⁰ *Alfarería, farmacia, ganadería, jardinería, manejo de caballos, minería, peinería, retórica, tonelería, vinatería, zoología.*

4. Artistic words in the two *Autoridades*

To provide an example of the use of the data contained in the portal, we will illustrate the analysis of entries for the Arts in the two *Autoridades*, that is, those of architecture, painting, and sculpture. The special attention we devote to architecture and painting is due to the fact that sculpture is represented much less, as can be seen in table 1.

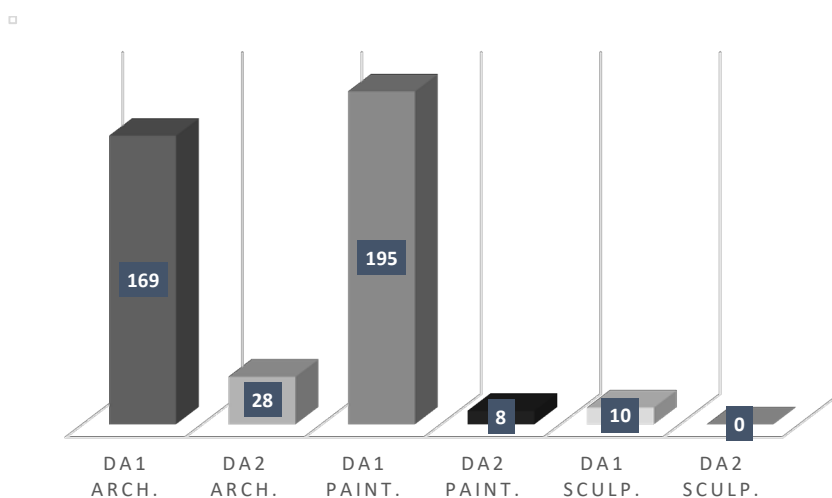


Table 1
DC (1768-88): Words relating to the Arts in DA1 and DA2.

4.1. Architectural and pictorial words sources

The sources of the entries for the Arts of DA1, as far as architecture is concerned, are many: technical texts include the *Tratado de Bóvedas* by Torija and the *Compendio Mathematico* by Tosca, whose book 5 is dedicated, among other subjects, to civil architecture, stonework, and military architecture. The *Museo Pictórico* of Palomino¹¹ is of particular importance.

Academics also cite legal documents such as the *Ordenanzas de Sevilla*, which contains technical terminology for masons and carpenters; they cite history books, such as *Antigüedades de España* by Bernardo de Aldrete and travel books like the one by Juan Cristóbal Calvete¹². The main lexicographical source is Covarrubias's *Thesoro*, mentioned several times in the initial paratexts. Not all the sources are cited in the only published volume of DA2: among those absent in the list that appears in the first edition of *Autoridades*, we find Gracián's *El Criticón*, the *Diálogo de Medallas* by

¹¹ He is also the author of the beautiful plate, engraved by his nephew Juan Antonio, which is among the paratexts of the *Diccionario de autoridades* of 1726.

¹² *Felicissimo viaje del Principe don Felipe* (Antwerp, 1552). This book deals with the trip to Flanders of Philip the Second.

Antonio Agustín¹³, the *Grandezas de la Villa de Madrid* by Jerónimo de Quintana, the *Casa de locos de amor* by Quevedo and the *Leon prodigioso* by Tejada, all of them used by academics as sources of authorities on architecture.

During the writing of DA2, the academics integrated the cited sources with more technical texts that are mainly contemporary: i.e., the *Compendio de los diez libros de arquitectura de Vitruvio* (1761), a translation by Joseph Castañeda from Claude Perrault's French text, the treaty *Elementos de toda la arquitectura Civil* (1763) by Christian Rieger, translated into Spanish by Benavente, and the *Arte y uso de la arquitectura: el primer libro de Euclides* whose Spanish translator is Lorenzo de San Nicolás (1796).

Among the works of the 16th century, the RAE also cites the translation of Villalpando of the *Tercero y cuarto libro de Architectura* by Sebastia[n] Serlio (1552), which is absent in DA1. These works are not found in the list included in the first volume of the second edition of *Autoridades*, corresponding to letters A and B.

The sources of words referring to painting in DA1 are also varied, but in this case, ninety percent of the words come from a technical work, the *Museo Pictórico* by Palomino, which is found again as the main source also in DA2.

4.2. Architectural words

In brief, not all the entries will be commented on, but there will be a focus on those on architecture in DA2. Among the words presented here, we have not considered the entries of DA1 that had only typographical changes in accents in DA2. The tables do not show the voices beginning with A and B published in 1770.

Most of DA2 words that refer to architecture designate column segments, ornamentations, and mouldings; these were included in the “common” dictionaries published while academics continued the work of writing the “large” dictionary. The cultisms *cariátide*, *caulícolo* and *cimacio*, the compound of probable Italian origin *contrapilastra*, as well as *cornijamento*, probably introduced from the Italian *corniciamento*¹⁴, entered the DRAE in 1780.

¹³ Except for the work of Gracián, we will find these sources in the DA2 list.

¹⁴ *Corniciamento* appears in the *Vita di Benedetto da Maiano. Scultore et architetto*, by Vasari (1568: 3, p. 530).

Entries	Sources
cariátide	Covarrubias, <i>Thesoro</i> (1611) Rieger, <i>Elementos de toda la arquitectura civil</i> . (1763)
caulículo	Tosca, <i>Compendio mathematico</i> . Tomo V (1712) Castañeda, <i>Compendio de Vitruvio</i> . (1761)
cimacio	Tosca, <i>Compendio mathematico</i> . Tomo V (1712)
contrapilastra	n/a
cornijamiento	Serlio, <i>Tercero y cuarto libro de Architectura</i> . (1552)

Table 2

Cultisms that enter the DRAE in 1780 and their sources.

Estría, *fastial*, *fusto*, the cultisms *denticulo*, *diástilo*, *embasamento*, *entrecolumnio*, the Italianism *escarzano*, *freso* – of disputed origin – the derivation *despiezo*, the compounds *engatillar* and *entrecanal* entered the DRAE in 1791. As in the DRAE of 1780, the sources are not from literary texts but technical ones:

Entries	Sources
denticulo, diástilo, engatillar, entrecanal, fusto	Castañeda, <i>Compendio de Vitruvio</i> . (1761)
despiezo, embasamento, entrecolumnio, fastial, freso	Diego de Sagredo, <i>Medidas del Romano</i> . (1526)
escarzano	Lorenzo de San Nicolás, <i>Arteanduso de la Arquitectura</i> . (1639 - 1665) [1736].
estría	Diego de Colmenares, <i>Historia de la Insigne Ciudad de Segovia</i> . (1637)

Table 3

Words that enter the DRAE of 1791 and their sources.

As far as neologisms are concerned, except for *anta* (Stevens 1706), *cariátide* (Vittori 1609 and Minsheu 1617), *diástilo* and *fastial* (Terreros 1786-1787) and *freso* (Sobrino 1705), the other entries appear for the first time in a lexicographical repertoire.

Table 4 shows the first textual testimony of the words: those written in regular characters are attested in the CDH, those in italic were found in

*Google books*¹⁵. Generally, they are well-established words in the language, except for *entrecanal*, *engatillar*, *fusto y gaucho*, whose first appearances are placed between 1761 and 1788.

Centuries	Words
XIII-XIV	fastial
XVI	cariatide, caulicolo, cimacio, <i>contrapilastra</i> , denticulo, diástilo, embasamento, escarzano, estría, freso, despiezo
XVII	cornijamento, <i>entrecolumnio</i>
XVIII	<i>entrecanal</i> , <i>engatillar</i> , fusto

Table 4
First textual testimony of the words.

5. Conclusions

As we have just seen, *Dicciocho. Portal lexicográfico del siglo XVIII* arises from the need to further our knowledge of the lexicographic documentation of the Age of Enlightenment and exploit it through its analysis. It is not so much a question of illustrating the results of our research, but rather that of providing a large amount of data on the works that constitute the basis of our research thus giving them a different life. We hope that our work will be, in turn, of service to other researchers who will be able to consider different aspects that we have not investigated.

We want to underline the need to consider the different editions of the academic dictionaries not as mere versions of the same work, but as the testimony of an evolution in the model of language reflected and in the lexicographic work, an evolution that reflects the mentality of the time. We do not only wish to illustrate the microstructural elements that characterize the lexicography of the eighteenth century: we would also like to contribute to the drawing up of a dictionary theory that stimulates the deepening of the networks of national and international relations between the different repertoires of the time.

Bionote: Elena Carpi is a specialist in lexicography and lexicology of the 18th century: she has analyzed the Terreros dictionary and the *Diccionario de autoridades*, both the first edition and the manuscripts of the second edition. She has been invited to join the Scientific Committee of the Centro Studi sul Settecento Spagnolo (CSSS) of the Alma

¹⁵ *Contrapilastra*: Lorenzo de San Nicolás (1639). *Entrecolumnio*, *entrecanal* and *engatillar*: (*Vitruvio*, 1761). *Gaucho*: Bails (1783).

Mater Studiorum Università di Bologna, and to join as an associate member the Feijoo Institute of 18th Century Studies (IFEES XVIII) of the University of Oviedo.

Author's address: elena.carpi@unipi.it

References

- Bails B. 1783, *Diccionario de Arquitectura Civil*, En la imprenta de la viuda de Ibarra, Madrid.
- Carpi E. 2021, *El léxico filosófico neológico en el Diccionario del padre Terreros y el Dictionnaire de Trévoux*, in Villar Díaz M.B., De Hoyos J.C., Dury P., Makri-Morel J., Renner V. (eds.), *La néologie des langues romanes: nouvelles approches, dynamiques et enjeux*, Berlin, Bern, Bruxelles, New York, Oxford, Warszawa, Wien, Peter Lang. pp. 61-77.
- Carpi E., Carriscondo Esquivel F. M. 2023, *Le Dictionnaire de Trévoux comme source du Diccionario Castellano con voces de Ciencias y artes (1786-1788): les nombres des cultes non catholiques*, in Galleron I. et Williams G. (éds.), *Dictionnaires et réseaux des lexicographes aux XVII^e et XVIII^e siècles*, Éditions Honoré Champion, Paris, pp. 183-202.
- Casas C. de las 1570, *Vocabulario de las dos lenguas toscana y castellana*, Francisco de Aguilar y Alonso Escribano, Sevilla.
- CDH: Real Academia Española 2013, *Corpus del Diccionario histórico de la lengua española (CDH)*, <https://apps.rae.es/CNDHE> (02.04.2023).
- Covarrubias S. de 1611, *Tesoro de la lengua castellana o española*, Luis Sánchez, Madrid.
- Franciosini L. 1620, *Vocabolario español-italiano, ahora nuevamente sacado a luz [...]*, Iuan Pablo Profilio, a costa de Iuan Ángel Rufineli, y Ángel Manni, Roma.
- Google books, <https://books.google.it/> (02. 04. 2023).
- Minsheu J. 1617, *Vocabularium Hispanicum Latinum et Anglicum copiosissimum, cum nonnullis vocum millibus locupletatum, ac cum Linguae Hispanica Etymologijs [...]*, Iohn Brownes, Londres.
- Oudin C. 1616, *Tesoro de las dos lenguas francesa y española. Thresor des deux langues françoise et espagnolle*, Marc Orry, París.
- Percivall R. 1599, *Bibliothecae Hispanicae pars altera. Containing a Dictionarie in Spanish, English and Latine*, John Jackson y Richard Watkins, Londres.
- Real Academia Española 1726-39, *Diccionario de la lengua castellana, en que se explica el verdadero sentido de las voces, su naturaleza y calidad, con las frases o modos de hablar, los proverbios o refranes, y otras cosas convenientes al uso de la lengua [Diccionario de autoridades]*, Francisco del Hierro (vols. I-II); Imprenta de la Real Academia Española, por la Viuda de Francisco del Hierro (vol. III); Imprenta de la Real Academia Española, por los Herederos de Francisco del Hierro (vols. IV-VI), Madrid.
- Real Academia Española 1770, *Diccionario de la lengua castellana compuesto por la Real Academia Española. Segunda impresión corregida y aumentada*. Tomo primero. A-B, Joaquín Ibarra, Madrid.
- Real Academia Española 1770-1829, *Manuscritos para la segunda edición del Diccionario de autoridades*. <https://www.rae.es/manuscritos-para-la-segunda-edicion-del-diccionario-de-autoridades> (2.04.2023).
- Sobrino F. 1705, *Diccionario nuevo de las lenguas española y francesa*, Francisco Foppens, Bruselas.
- Stevens J. 1706, *A new Spanish and English Dictionary. Collected from the Best Spanish Authors Both Ancient and Modern [...]. To which is added a Copious English and Spanish Dictionary [...]*, George Sawbridge, Londres.
- Terreros y Pando E. de 1987 [1786-93], *Diccionario castellano con las voces de ciencias y artes*, ed. facs., 4 vols., Arco Libros, Madrid.

- Vasari G. 1568, *Le vite de' più eccellenti pittori, scultori e architettori*, Giunti, Firenze.
Consulted in *Giorgio Vasari Le Vite-Edizioni Giuntina e Torrentiniana*,
<http://vasari.sns.it> (02. 04. 2023).
- Vittori G. 1609, *Tesoro de las tres lenguas francesa, italiana y española. Thresor des trois langues françoise, italienne et espagnolle*, Philippe Albert & Alexandre Pernet, Geneve.

Annex 1

agricultura	curtiduría	maestros de coches
alarifes	danza	mampostería
albañilería	derecho	maquinaria
albeitería	dialéctica	marinería
álgebra	dióptrica	matemática
alquimia	doradores	medicina
altanería	ebanistería	metafísica
altareros	ejercicio de caballos o mulas	milicia
anatomía	empedrados	mitología
aritmética	ensambladores	monedería
armería	escolástica	montería
arquitectura	escribanía	música
arsenales	escuelas	náutica
artillería	escultura	navíos
astrología	esgrima	notaría
astronomía	espagírica	oficiales
ballestería	estática	óptica
blasón	facultativas	panadería
bordadores	farmacopea	pañería
botánica	filosofía	perspectiva
boteros	física	pintura
botica	fontanería	platería
cabestreros	lo forense	polvoristas
cancelaría	fortificación	portaventaneros
cantería	fundición	química
carpintería	ganado	sederos
catóptrica	geografía	soladores
caza	geometría	soldadores
cerería	gnomónica	sombrerería
cerrajería	gramática	teatro
cetrería	guarnicionería	tejeduría
cirugía	herradores	tesorería
cocheros	hidrometría	tintorería
cocina	imprensa	toreo
colegios y universidades	ingenios	tribunales
comercio	juegos	trigonometría
contaduría	jurisprudencia	vaciadores
contratos	labranza	volatería
cordelería	lapidarios	zapatería
costura y sastrería	lavanderas	zurRADORES
cronología	libreros	
curia	lógica	

Annex 2

Number	Principle	End	Years
1	c	cebada	1780
2	cebada	confundirse	1780
3	confusamente	czarina	1780
4	d	disyunto	1789
5	disyunto	ezquierda	1789
6	f	horror	1789
7	horror	lynce	1789-1791
8	m	nefandísimo	1803-1817
9	paredaño	prosperidad	1803-1817

Table 5
Folders to *prosperidad*.

Number	Principle	End	Years
1	punar	quedito	1818
2	puño	quedito	1818
3	quedo	quizá o quizás	1818
4	r	rapaz	1818-1828
5	rapaz	re	1818-1828
6	re	receta	1818-1828
7	receta	rechoncho	1818-1828
8	rechoncho	reforma	1828
9	reforma	registro	1828-1829
10	registro	rejo	1829
11	rejo	remate	1829
12	remate	renovador	1829
13	renovador	repicar	1829
14	repicar	rescatar	1829
15	rescatar	restitución	1829
16	restitución	retrocesión	1827
17	retrocesión	reyerta	1829
??	retrocesión	rey	
18	reyerta	robo	1829
19	robo	rompedera	1829
20	rompedera	rotar	1829
21	saciedad	salaz	1829
22	salaz	saludable	1829
23	saludable	sanidad	1829
24	sanidad	secuencia	1829
25	secuencia	semidoble	1829
26	semidoble	serpiente	1829
27	serpiente	silvestre	1829
28	solfa	sordamente	1829

Table 6
Logbook from *punar* to *sordamente*.

ALGUNAS CALAS EN LOS DIEZ PRIMEROS CAPÍTULOS DE LA PRIMERA PARTE DEL *DON CHISCIOTTE DE FRANCIOSINI* “Toscanismos”, religión y germanía

JOSÉ FRANCISCO MEDINA MONTERO
UNIVERSITÀ DI TRIESTE

Abstract – Among the numerous translators into Spanish of the 17th century, the figure of the Italian Lorenzo Franciosini stands out. The production of this grammarian, lexicographer, and author of the first translation into Italian of *The Ingenious Gentleman Don Quixote of La Mancha* spans from 1620 to 1644 and includes a total of eleven works. In relation to his translation work, this paper aims to compare a small part of the original text of Cervantes’ *Don Quixote* (specifically the first ten chapters of the first part) with its translation (recall that he translated the first part in 1622 and the second in 1625). In these ten chapters, we have decided to pay attention to three very interesting aspects: “toscanismos” (Tuscanisms), religion, and “germanía” (a type of slang). We have focused on them because they often did not receive the appropriate translation by Franciosini, nor have they enjoyed sufficient interest in terms of linguistic and translator analysis. Likewise, their examination will serve us to have another useful parameter when evaluating the quality of the translation by our Tuscan author. In general terms, his version is reliable, although we have found numerous errors resulting, among other things, from its excessive literalness and the fact that our translator failed to understand certain passages of the original text.

Keywords: *Quixote*; Franciosini; Tuscanisms; religion; slang.

1. Introducción

La importancia que representa Cervantes para la literatura universal y para la cultura europea [por ejemplo para la pintura (Robert Combas, Hervé di Rosa...), para la música (Jakob Ludwig Felix Mendelssohn Bartholdy, Richard Georg Strauss...) o para el teatro (Jerónimo López Mozo, Charles Dibdin...)] resulta indiscutible. Gran parte de esa relevancia se debe a la publicación de *El ingenioso hidalgo don Quijote de la Mancha*, una novela

moderna¹ que influyó sobremanera en la narrativa europea de los siglos sucesivos.²

Una de las claves del éxito de este texto radica en que no solo se concibió como la sátira punzante y divertida de un género novelesco determinado, los libros de caballerías, que contaba con una enorme tradición en la época, sino también como una mofa muy irónica de todo el sistema social del momento. Como se sabe, en los libros de caballerías la acción se antepone a los personajes, algo que Cervantes contribuyó a modificar, al caracterizar psicológicamente a la perfección a los protagonistas de su obra.

Otra característica que contribuyó a su éxito descansa en su polifonía, ya que el alcaíno interpretó la realidad del momento desde distintas perspectivas. Esto ayudó a que el *Quijote* se convirtiera en la primera novela moderna. Debido a la existencia de varios narradores diferentes, los aspectos lingüísticos³ (registros de lengua, dobles sentidos, etc.) cobran un interés especial. Sin embargo, como veremos más adelante, esta riqueza lingüística genera no pocas dificultades cuando se traduce a lenguas como el italiano. No estamos refiriéndonos al registro lingüístico culto, en el que abundaban menciones al ámbito de la filosofía, de la literatura o del derecho (entre otros), sino al registro coloquial⁴ (y, dentro de él, al uso de la lengua espontánea, sencilla y natural de algunos de sus personajes). Esta circunstancia provocó que en numerosas ocasiones Lorenzo Franciosini (lexicógrafo, gramático y traductor), el primero que vertió la obra de Cervantes al italiano, no consiguiera plasmar con acierto algunas expresiones de la obra cervantina.

Con respecto a sus traslaciones, suele coincidirse en que el *Quijote* se encuentra entre los textos más traducidos de todos los tiempos. A este grupo tan selecto pertenecen, por ejemplo, la *Biblia*, *El Principito* (de Antoine de Saint-Exupéry), *Las aventuras de Pinocho* (de Carlo Collodi) o *Las aventuras de Alicia en el país de las maravillas* (de Lewis Carroll). Si acudimos al Instituto Cervantes, en concreto a la página web consagrada al “Mapa mundial de la traducción”,⁵ leemos que nuestra obra maestra se ha trasladado a más de 145 lenguas, dialectos o variedades lingüísticas. Si no erramos, la última traducción de este texto se realizó al patsuezu, la variante

¹ Vid. sobre el *Quijote* y la novela moderna, verbigracia, Elizalde Armendáriz (1981, pp. 949-962).

² Consúltese en relación a la ingente bibliografía que ha generado el *Quijote* la excelente recopilación de voces bibliográficas de Fernández S. J. (2008). Ahí surgen las citas de miles de artículos y libros dedicados a asuntos como, por ejemplo, el *Quijote* y la música, la astronomía, la sátira, la educación, la medicina, la gastronomía, la religión, el derecho... Más en concreto, si se desea acceder a la bibliografía referente al perfil del cervantismo italiano hasta 1991, conviene recurrir a Pini Moro y Moro (1992, pp. 149-268).

³ Véase sobre la lengua de Sancho y Don Quijote, por ejemplo, Garatea Grau (2007, pp. 171-186).

⁴ A nuestro juicio, este resulta más abundante en la primera parte del *Quijote* que en la segunda.

⁵ <https://mapadelatraduccion.cervantes.es/obras>

occidental del asturleonés, en 2023. Este hecho indica que el *Quijote* continúa suscitando un gran interés, al considerarse una obra muy moderna e incluso actual.

Ya en relación a nuestro artículo, hemos de señalar que a lo largo de las próximas páginas intentaremos profundizar en las propuestas traductorales que Franciosini otorgó a los “toscanismos” y a las lexías pertenecientes a la religión y a la germanía que se insertan en los diez primeros capítulos de la primera parte del *Quijote* cervantino (recuérdese que esta consta de 52 capítulos y la segunda de 74). Nos hemos centrado en estos aspectos porque creemos que a veces Franciosini se equivocó al verterlos y porque tenemos conocimiento de que hasta el momento han recibido poco interés lingüístico y traductor. Reputamos que este estudio y los trabajos que hemos publicado con anterioridad sobre la traducción de Franciosini del *Quijote* (Medina Montero 2015, pp. 203-224; 2016, pp. 147-159; 2017, pp. 171-186; 2018a, pp. 9-22; 2018b, pp. 25-39; 2019, pp. 61-78) nos servirán para crearnos una opinión inicial relativa a la calidad de la primera traslación de esta obra al italiano.

Por lo que atañe a la metodología, primero hemos comparado, palabra por palabra, los diez primeros capítulos de la primera parte de las versiones originales de Cervantes (1605) y Franciosini (1622). Después hemos seleccionado y hemos analizado las piezas léxicas inherentes a las tres esferas precedentes. Para estar seguros de su significado y de su pertenencia a ellas nos hemos servido de dos diccionarios de la época, el de Covarrubias y Orozco (1611) y el de la Academia de la Crusca (1612). Asimismo, hemos utilizado otros dos repertorios bibliográficos algo más modernos (sobre todo el italiano), el *Diccionario de Autoridades* (1726-1739) y el *Dizionario della Lingua Italiana* de Battaglia (1964). Los dos últimos han resultado cruciales para, desde la distancia cronológica, confirmar la veracidad de las acepciones de los vocablos localizados en los dos primeros diccionarios. Después hemos intentado indagar si Franciosini los había traducido bien o no y hemos tratado de exponer y justificar nuestros pareceres al respecto.

Hemos optado por dedicarnos al examen del léxico de la germanía porque, a nuestro juicio, su traslación a cualquier idioma entraña enormes dificultades. Por su parte, el estudio de los “toscanismos” se debe a nuestro deseo de conocer a qué ámbitos pertenecían y en qué contextos y por qué se emplearon. También hemos decidido llevar a cabo un análisis del léxico religioso, porque nos ha llamado la atención el hecho de que estos elementos desaparecieron casi por completo de la versión de Franciosini. Pese a todo, nuestra intención no finaliza aquí, ya que lo que de verdad hemos pretendido ha sido iniciar a reflexionar sobre cómo trasladó nuestro gramático al italiano este texto cervantino.

Por lo que concierne a la estructura del artículo, tras esta introducción tendrán cabida un capítulo consagrado a la figura de Lorenzo Franciosini,

otro a las traducciones del *Quijote* al italiano, tres en los que nos ocuparemos de los tres campos que acabamos de mencionar (los “toscanismos” y los términos relacionados con la religión y la germanía) y un último apartado en el que efectuaremos las conclusiones sobre nuestro trabajo.

2. Lorenzo Franciosini

Pese a haber consultado abundante material bibliográfico, hemos hallado pocos datos sobre la biografía de Lorenzo Franciosini, a quien la crítica sitúa entre los mejores hispanistas extranjeros del siglo XVII. En aquella época, a veces las noticias relativas a los escritores surgían en las portadas de sus textos.⁶ Sirvan de ejemplo los casos de Francisco de Quevedo Villegas, “cauallero del Orden de Santiago”, de sor Juana Inés de la Cruz, “monja professa en el monasterio del Señor San Gerónimo de la ciudad de México”, o de Alonso de Ledesma, “natural de Segouia”.

Por lo que se refiere a nuestro autor, en las portadas de su *Vocabolario italiano, e spagnolo* (1620), de su traducción del *Quijote* (1622 y 1625) y de su *Gramatica spagnola, e italiana* (1624) se lee que es “fiorentino”. En cambio, en la de sus *Dialogos apazibles, compvestos en Castellano, y traduzidos en Toscano* (1626) se va más allá y se especifica que nació en Castelfiorentino (provincia de Florencia) y que, además, dio clases de italiano y español en Siena.

Del Bravo aporta algunas noticias más sobre la vida de Franciosini, que aparecen en las fuentes a las que ha recurrido y que ha citado. Así, en un artículo publicado en el *Dizionario Biografico* Treccani, consultado en 2024 [[https://www.treccani.it/enciclopedia/lorenzo-franciosini_\(Dizionario-Biografico\)/?search=FRANCIOSINI%2C%20Lorenzo](https://www.treccani.it/enciclopedia/lorenzo-franciosini_(Dizionario-Biografico)/?search=FRANCIOSINI%2C%20Lorenzo)], afirma que nuestro autor vivió en Florencia en 1637 y que seguramente murió después de 1645. Muy poco más conocemos de su vida.

La situación difiere por lo que respecta a su producción bibliográfica, ya que de esto se sabe mucho más. De sus once obras, compuestas entre 1620 y 1644 (adviértase que entre 1620 y 1627 vieron la luz nada más y nada menos que seis de ellas), destacan el *Vocabolario italiano, e spagnolo*⁷ (Roma, 1620) y la *Gramatica spagnola, e italiana* (Venecia, 1624), dos textos destinados a la enseñanza del español a los italianos que se emplearon hasta, como poco, finales del siglo XVIII. En las primeras páginas de su *Vocabolario italiano, e spagnolo* Franciosini inserta una pequeña “Introduzione alla lingua Spagnola”, un prelude de las cuestiones que, más tarde, encerrará su *Gramatica spagnola, e italiana*. Con respecto a la creación

⁶ O en las notas preliminares de los mismos.

⁷ Léase en relación a este vocabulario, Martínez Egido (2002).

de esta obra, la crítica coincide en sostener que Franciosini se sirvió de la *Grammaire espagnolle explicqvee en Francois* de César Oudin⁸ (1597) y que, a su vez, este no perdió de vista las *Osservationi della lingua castigliana* de M. Giovanni Miranda (1566).

Franciosini completó la segunda edición de su *Gramatica spagnola, e italiana* (Roma, 1638) con tres libros prácticos publicados antes (en 1626) en la ciudad de los canales, los *Dialogos apazibles, compvestos en Castellano, y traduzidos en Toscàno*, los *Dichos Politicos, y Morales* y su *Nomenclator, ò Registro de algunas cosas curiosas, y necessarias de saberse à los estudiosos de la lengua Española*. El primero se basó en los *Dialogos muy apazibles, escritos en lengua Española, y traduzidos en Frances* de César Oudin (1608), quien tomó como modelo los *Pleasant and Delightfvl Dialogues...* del inglés John Minsheu (1599). Los *Dichos Politicos, y Morales* surgen en dos columnas, la primera en español y la segunda en italiano. Cuando Franciosini hubo de localizar en italiano las correspondencias de esas estructuras fijas españolas, en no pocas ocasiones realizó traslaciones muy literales y, por ende, carentes de sentido, una constante en su labor de traductor. Por último, su *Nomenclator* consiste en la presentación de una nomenclatura bilingüe italiana y española organizada por capítulos, en los que el léxico aparece por unidades temáticas (las partes del cuerpo humano, los meses del año, los vestidos masculinos y femeninos, etc.).

En su faceta de traductor, Franciosini vertió a nuestra lengua *El ingenioso hidalgo don Quijote de la Mancha* con el título de *L'ingegnoso Cittadino Don Chisciotte della Mancia. Composto da Michel di Cervantes Saavedra. Et hora nuouamente tradotto con fedeltà, e chiarezza, di Spagnuolo, in Italiano. Da Lorenzo Franciosini fiorentino*. La traducción la imprimió Andrea Baba en Venecia en dos momentos diferentes, en 1622 (la primera parte, que consta de 669 páginas) y en 1625 (la segunda, que contiene 751 páginas). Franciosini le dedicó la primera parte al Gran Duque de Toscana, “Don Ferdinando Secondo”, que nació en 1610 y falleció en 1670, y la segunda al señor “Ferdinando Seracinelli (sic), Balì di Volterra” (1583-1640).

Su producción incluye un libro ideado para que los extranjeros aprendieran nuestro idioma con más facilidad, mediante una serie de lecturas complementarias en clave de humor. Lo tituló *Rodomontadas españolas, Recopiladas de los Comentarios de los muy espantosos, terribles e inuencibles Capitanes, Matamoros, Crocodilo, y Rajabroqueles* (Venecia, 1627).

⁸ De Riquer y Morera (1942, p. 21) describe una diferencia crucial entre César Oudin y Lorenzo Franciosini: “Franciosini siente un fervoroso amor a España, por su lengua y por su literatura; Oudin nos profesó una cordial antipatía, y su labor de hispanista le fué impuesta por su condición de intérprete real”.

Franciosini también se ocupó de la enseñanza de la lengua italiana. Para ello escribió dos textos editados en Florencia, el primero en 1637, *De particvlis Italicae Orationis, qvibus accessit...*, en cuya portada se lee que aquel año vivía en aquella ciudad, y el segundo, su *Fax lingvae italicae*, en 1638.

Por último, San Vicente Santiago (2016, p. 1) atribuye otros dos libros a Franciosini, la “*Introduttione alla lettura et intelligenza grammaticale della lingua spagnuola* (Siena, 1644), obra que fue publicada tanto por separado como incluida en la conocida gramática trilingüe de Lonchamps”, y “una reducción del *Fax* [...], el *Compendium facis linguae Italicae* (Siena, 1644)”, un resumen de la edición de 1638 en la que se contemplan, en relación a esta, la misma dedicatoria al lector, pero también algunas variaciones.

3. Las traducciones del *Quijote* al italiano

No cabe duda de que la traducción (escrita u oral) siempre ha desempeñado un papel crucial en nuestras vidas. Sin ella, las barreras idiomáticas entre las personas se mantendrían y los avances tecnológicos, científicos, etc. sufrirían una enorme ralentización. En este escenario imaginario también se vería afectada la transmisión del saber cultural, ámbito en el que se incluye la traslación de obras como el *Quijote*.

Las claves del éxito de esta novela se basan, verbigracia, en su carácter innovador y en la calidad que atesora, algo irrefutable. No obstante, ha de reconocerse que su traducción a tal número de lenguas, dialectos o variedades lingüísticas⁹ también ha ayudado a que el texto culmen de la literatura española forme parte del selecto grupo de obras más importantes de la literatura universal de todos los tiempos.

Por lo que respecta a los idiomas de nuestro entorno cultural más cercano, el *Quijote* se vertió primero al inglés. Lo hizo Thomas Shelton en 1612 (la primera parte) y en 1620 (la segunda parte). Después se trasladó al francés, gracias a César Oudin (la primera parte data de 1614) y a François de Rosset (la segunda se efectuó en 1618), y muy poco más tarde al italiano (Lorenzo Franciosini se encargó de la primera parte en 1622 y de la segunda en 1625). Y por lo que atañe al alemán y al neerlandés, Pahsch Basteln von der Sohle, seudónimo de Joachim Caesar, tradujo en 1648 los primeros

⁹ Nos parece harto interesante el comentario de Scaramuzza Vidoni (2013, p. 221), quien asevera que las traducciones y adaptaciones del *Quijote* nunca han prescindido del contexto cultural de los distintos momentos en los que se han llevado a cabo:

Se ci concentriamo sul caso illustre del capolavoro di Cervantes troviamo un fiorire di traduzioni e adattamenti che non sono semplicemente calchi o riprese, ma rielaborazioni e trasformazioni legate agli ambienti delle varie epoche e ai processi di mediazione del contesto culturale.

veintidós capítulos de la primera parte al alemán y Lambert van den Bosch (también Lambert van den Bos o L. Sylvius) vertió la obra completa al neerlandés en 1657.¹⁰

Por lo que concierne a las traducciones del *Quijote* al italiano, para comenzar deseamos destacar que merced a su difusión por el país de la Bota, Cervantes conquistó allí un mayor reconocimiento. Como consecuencia, en 1626 se trasladaron en Venecia otros dos textos suyos, a saber, las *Novelas exemplares* (*Il Novelliere Castigliano*, labor de Gvglielmo Alessandro de Nouilieri, Clauelli) y *Los trabajos de Persiles, y Sigismvnda, Historia Setentrional* (*Istoria settentrionale, De trauagli di Persile, e Sigismonda*, versión de Francesco Ellio Milanese).

A partir de 1625, durante casi los dos siglos sucesivos no se realizó ninguna traslación¹¹ del *Quijote* al italiano. Así las cosas, hemos de preguntarnos si esta circunstancia se debió a la enorme calidad que escondía la traducción de Franciosini (lo cual habría contribuido a que perdurara tanto tiempo) o si, en cambio, este hecho dependió de la transmisión que alcanzó la novela en Italia, en comparación con la que logró en países como Francia, Inglaterra o Alemania, razón por la cual no habría interesado verterla, o de otros factores. Analicemos con brevedad algunas cuestiones.

Por un lado, por lo que se refiere a su calidad, si escudriñamos el texto de Franciosini nos daremos cuenta de que nos enfrentamos a un producto literario que no respeta uno de los objetivos primordiales de Cervantes, esto es, el acometimiento de una sátira irónica y humorística contra los libros de caballerías. La responsabilidad recae en su traducción, que transmite una excesiva comicidad que persigue, ante todo, la diversión y el deleite del lector, aspecto este (la diversión desmesurada, sin más) ausente en la obra cervantina. Y para más inri, a veces determinados fragmentos resultan muy literales e incluso incomprensibles, algo en lo que incidiremos a lo largo de las próximas páginas.

Por otro lado, conviene recordar que aunque el *Quijote* tardó en propagarse por Italia en los siglos XVII y XVIII, consiguió influir algo en la literatura de aquel país. Pero lo hizo presentando una clave de lectura

¹⁰ Si se desea profundizar en estas cuestiones, Colón i Domènech (1974) compara, si bien de forma somera, las primeras traslaciones del *Quijote* al italiano, al francés, al alemán, al inglés y al neerlandés.

En referencia a algunos territorios en los que se hablaba este último idioma, Luttikhuizen (2008, p. 306) aclara sobre la relación entre los textos de Cervantes y los Países Bajos que “(...) el interés por las obras de Cervantes se manifestó en los Países Bajos de una manera totalmente distinta a la de otros países”. En efecto, este inició gracias a una adaptación poética al neerlandés del poeta y jurista Jacob Cats (1577-1660) de *La Gitanilla* (Dordrecht, 1637), el relato breve que abre la colección de las *Novelas exemplares* cervantinas.

¹¹ También hubo pocas reediciones y reimpressiones, en comparación con lo que ocurrió en Francia, Inglaterra y Alemania, asunto en el que no vamos a inquirir en estas páginas.

inapropiada con respecto a la intención de Cervantes, esto es, se difundió, con la misma comicidad desmedida que se observa en la traslación de Franciosini, en dramaturgos como Girolamo Gigli, con *Ludovico Pio* (1687), *Un pazzo guarisce l'altro* (1687), *Atalipa* (1689) y *Amore fra gli impossibili* (1693), o como Marco Morosini con su *Il don Chissiot della Mancia* (1680), como recoge su portada, un “Drama per mvica”.

Asimismo, en nuestra opinión, otra de las causas de la escasez de traslaciones al italiano durante esos casi doscientos años se asienta en que los lectores cultos de aquellos siglos poseían la capacidad suficiente para leer el texto directamente en español.

Tras la traducción de nuestro lexicógrafo le tocó el turno, casi dos siglos después (en concreto en 1818), a la de Bartolomeo Gamba. Gamba, que nació en Bassano del Grappa (Vicenza) en 1766 y murió en Venecia en 1841, utilizó para su traducción, que se imprimió en Milán (Andrea Ubicini), la edición de 1780 de la Real Academia Española. Esta, fruto del trabajo del impresor y encuadernador zaragozano Joaquín Ibarra Marín, constaba de cuatro tomos y encerraba treinta y seis magníficos grabados. Para concluir este apartado, hemos de señalar que en un estudio reciente (2019, p. 88) Botta afirma que en el siglo XX se publicaron once traslaciones al italiano y que, hasta ahora, en el siglo XXI han visto la luz otras cuatro.

4. Los “toscanismos”¹²

Franciosini recurrió a veces al léxico toscano porque procedía de allí. Si hubiera nacido, por ejemplo, en Cerdeña, tal vez habría empleado lexías del sardo. Esto sucedió porque, desde nuestro punto de vista, quienes conocen la fragmentación lingüística existente en Italia (acentuada aun hoy en día) saben que mediante los dialectos pueden expresarse con más eficacia y mayor vivacidad que a través del italiano determinadas expresiones que pertenecen sobre todo a la oralidad y al lenguaje coloquial.

Por medio de esta solución, nuestro lexicógrafo intentó resolver el problema que le crearon algunos pasajes o estructuras lingüísticas de la novela de Cervantes que rebosan espontaneidad (destaquemos, verbigracia, las fórmulas expresivas). Franciosini también acudió a su dialecto para apropiarse al lector toscano, cuando hubo de trasladar culturemas referidos a instituciones, a medidas de capacidad, etc., cuando tuvo que

¹² En los ejemplos que expondremos sobre el léxico toscano, religioso y de la germanía aparecerán entre paréntesis solo las fuentes lexicográficas en las que hemos hallado las expresiones en cuestión. Recordemos los repertorios bibliográficos consultados: dos diccionarios de la época, el de Covarrubias y Orozco (1611) y el de la Academia de la Crusca (1612), y otros dos algo más modernos (en especial el italiano), el *Diccionario de Autoridades* (1726-1739) y el *Dizionario della Lingua Italiana* de Battaglia (1964).

aclarar algunos dobles sentidos cervantinos o cuando pretendió que se comprendiera con más claridad el texto, para lo cual incluso optó por la realización de ampliaciones lingüísticas, un fenómeno cada vez más frecuente a medida que va avanzando su traducción. A continuación analizaremos todos los “toscanismos” que hemos localizado en nuestro corpus.

Por lo que atañe a la versión de los culturemas, el primer contratiempo aflora en el mismo título del libro, ya que **hidalgo** (1605, f. 1 r, Covarrubias y *Autoridades*) se transforma en *cittadino* (1622, p. 1, Crusca y Battaglia).¹³ Al haber elucidado con detenimiento los motivos de esta elección en un trabajo anterior (Medina Montero 2015, pp. 214-216), al que remitimos, en las próximas líneas nos limitaremos a resumir con brevedad lo que planteamos allí.

En palabras de Vega Cernuda (2005, p. 13), “ni *gentleman*, ni *gentilhomme*, ni *nobiluomo* (ni, a nuestro juicio, *cittadino*) recogen el contenido semántico de ese hijodalgo, típicamente hispano, que supone una autoestima personal más que una condición social”, dictamen coincidente con lo que contiene el lema **fidalgo** de Covarrubias: “este termino es muy propio de España”. *Ibidem* se lee que **fidalgo** “Equiuale a noble, castizo, y de antigüedad de linage: y el ser hijo de algo, significa auer heredado de sus padres y mayores, lo que llama algo, que es la nobleza”. Por su parte, en *Autoridades* **hidalgo** se describe como “La persona noble que viene de casa y solar conocido, y como tal está exento de los pechos y derechos que pagan los villanos”. De estas definiciones se infiere que los hidalgos formaban parte de la nobleza y que, aunque en aquel período este estamento había perdido parte de sus beneficios, como tal aún poseían ciertos privilegios. Veamos ahora qué ocurre con el término *cittadino*.

Franciosini definió las piezas léxicas **hidalgo** y *cittadino* tanto en su *Vocabolario italiano, e spagnolo* de 1620, como en su *Gramatica spagnola, e italiana* de 1624. Según él, **hidalgo** “significa propriamente Cittadino” (1624, p. 262), es decir, un hombre “ben nato, e privilegiato da certi dazj, o gabelle” (1620, p. 301 de la segunda parte). Asimismo, por *cittadino* entendía una “persona ben nata, e capace d’alcuni honori, e privilegij del suo paese” (1624, p. 262) que “è capace de gli honori, e de benefici della città” (1620, p. 164 de la primera parte), conceptos que aparecieron años antes (en 1612) en el *Vocabolario degli Accademici della Crusca* (persona “che è capace degli onori, e de’ benefici della Città”). Todo se mueve, entonces, en torno al asunto de los privilegios o de los beneficios de los que gozaba el *cittadino* toscano (y el **hidalgo** español), pero de los que carecía el *cittadino* del resto

¹³ Franciosini no escogió el préstamo del español *idalgo*, según Battaglia presente en italiano desde la obra de 1565 de Cristoforo Montanini titulada *Gli effetti d’amore*, a causa de su escasa frecuencia en los textos de aquellos siglos.

del país de la Bota, ya que ninguna de las acepciones anteriores (toscanas) del siglo XVII se encuentra en el repertorio bibliográfico de Battaglia.

Otros cultuemas han corrido igual suerte. Así, Franciosini se sirvió del dialecto toscano para traducir **galgo** (1605, f. 1 r, Covarrubias y *Autoridades*) con *can d'aggiugnere* (1622, p. 1), una expresión que solo se da cita en Crusca, razón por la cual reputamos que pertenece al caudal léxico del toscano, y la pieza léxica **hanega** (1605, f. 1 v, Covarrubias y *Autoridades*) con el término toscano *staiuòro* (1622, p. 2, Battaglia).

Nuestro gramático también prefirió aproximar al universo toscano dos instituciones de nuestro entorno cultural, el **Iurado de Cordoua** y la **santa Hermandad**. Para la primera (1605, f. 22 r) Franciosini eligió *quarantotto* (1622, p. 50; *quarantotto/-i*, Crusca, pero la edición de 1691, y Battaglia). Ambas corporaciones formaban parte de las antiguas élites de poder político de las ciudades de, respectivamente, Córdoba y Florencia. Franciosini pensó en los *Quarantotti* porque esos cuarenta y ocho senadores florentinos ejercían funciones similares a las de los miembros del jurado de Córdoba. Por lo que, en cambio, concierne a **santa Hermandad** (1605, f. 35 r, Covarrubias y *Autoridades*), un tribunal que contaba con plena jurisdicción para castigar “los delitos cometidos en el campo” (Covarrubias y *Autoridades*), nuestro lexicógrafo propuso *Bargel di Campagna* (1622, p. 78, Battaglia), una colocación empleada, según Battaglia, “specialmente a Firenze”. Pese a este intento, Franciosini no acertó de pleno porque el *Bargel di Campagna* actuaba en cualquier contexto y no solo en el campo, como ocurría con la **santa Hermandad**.

De igual modo, Franciosini se valió del toscano para trasladar ciertas fórmulas expresivas. Así, para **a la buen hora** (1605, f. 11 v), según Rico Manrique (1998) “en buena hora”, seleccionó *con cento buon'anni*¹⁴ (1622, p. 25), una unidad pluriverbal que solo hemos localizado en el *Thesoro della lingua toscana* de Da Montemerlo (1594, p. 298). Por último, para verter la forma de juramento **Viue Roque** (1605, f. 13 v) nuestro lexicógrafo se inclinó por *al corpo di Sampuccino* (1622, p. 30).

Reflexionemos con laconismo sobre esta estructura. Tras haber consultado abundante material bibliográfico, nos atrevemos a conjeturar que quizás ese *Sampuccino*, inserto en la segunda parte de la fórmula exclamativa, traiga a colación el personaje de Fra Puccino, un fraile que algo menos de tres siglos antes de 1622 se relacionó con Castelfiorentino, la población en la que, como se lee en la portada de los *Dialogos apazibles* de 1626, nació Franciosini. He aquí las palabras de Repetti extraídas de su

¹⁴ A nuestro parecer, la expresión *con mille buon'anni* (1622, p. 48), la traducción de Franciosini de **en ora buena** (1605, f. 21 v, *Autoridades*, pero bajo **enhorabuena**), también podría provenir del toscano, porque semeja sobremanera a *con cento buon'anni*.

Supplemento al Dizionario geografico fisico storico della Toscana (1845, p. 59) en las que se menciona ese vínculo:

CASTEL-FIORENTINO. - Dove si discorre del ponte sull’Elsa a Castel Fiorentino si aggiunga: che fu riedificato nel 1354 a spese della Mansione dell’Altopascio, siccome lo da a conoscere una provvisione del 26 giugno di detto anno fatta in Firenze dai Signori della *Parte* deputati alle strade e ponti, quando fu nominato Fra Puccino converso della Badia di Vallombrosa in esattore delle rendite della Mansione de’Fрати dell’Altopascio obbligati alla riedificazione del Ponte di Castel Fiorentino. - (ARCH. DIPL. FIOR. *Carte della Vallombrosa*)

En su traducción, Franciosini efectuó a veces ampliaciones lingüísticas. Mediante esa operación pretendía hiperbolizar el escrito original, al que poco a poco fue despojando de su carácter primitivo de parodia de los libros de caballerías y fue transformando en un texto un tanto diverso, en una variante cómica de dichos libros, muy del gusto itálico de la época. Nuestro lexicógrafo, pues, optó por añadir a lo que halló en la obra cervantina elementos como, verbigracia, adverbios [**Tirad** (1605, f. 10 v, Covarrubias y *Autoridades*) > *tirate pure allegramente* (1622, p. 23, Crusca y Battaglia)], verbos [**enuistio** (1605, f. 26 v, Covarrubias y *Autoridades*) > *assaltò, & investì* (1622, p. 59, Crusca y Battaglia)] o sustantivos [**despojos** (1605, ff. 25 v y 29 r, Covarrubias y *Autoridades*) > *preda, & bottino* (1622, pp. 58 y 65, Crusca y Battaglia)]. A esta última categoría morfológica pertenece el vocablo *crazia*, una moneda toscana que Franciosini citó para hacer un guiño a sus lectores de aquella zona. Pero al mismo tiempo, con fines clarificadores, le situó al lado otra unidad monetaria, el *baiocco*, por aquel entonces más habitual en el resto de Italia. Estas dos piezas léxicas, *crazie, ò baiocchi* (1622, p. 6, ambas solo en Battaglia), traducen la moneda española **real** (1605, f. 3 r, Covarrubias y *Autoridades*).

Por último, para intentar otorgarle el doble sentido, jocosos, a una de las denominaciones que recibió el protagonista de nuestra historia, **Quixada**¹⁵ (1605, f. 1 v, Covarrubias y *Autoridades*), vinculada con el rostro de don Quijote, nuestro gramático escogió un término agrícola toscano de la época, *Chisciàda* (1622, p. 2, Battaglia), que, de todos modos, no se adscribe al mismo ámbito (el somático) de **Quixada**. Battaglia atribuye el origen de *Chisciàda* a *chisciare*, que para Crusca, que remite a *sarchiellare*, significa “Leggiermente sarchiare”.

¹⁵ Esto no lo logró ni con **Quixote** (*Chisciotte*), ni con **Quesada** (*Chesàda*).

5. El léxico religioso

En palabras de Bernardi (1995, p. 96), “[...] el escamoteo de la alusión religiosa mediante el empleo del inofensivo término **médico** es un recurso que Franciosini utiliza otra vez a lo largo de la traducción”. Bernardi está refiriéndose a la traslación por parte de Franciosini de, entre otras, la pieza léxica **patriarca** como *Medico*. En los capítulos examinados hemos notado que nuestro gramático tradujo casi todos los vocablos religiosos mediante lexías referentes a campos semánticos diferentes al de la religión. Esto sucedió cuando habló de Dios, del Evangelio, del diablo, de figuras más terrenales como los frailes, los religiosos, etc.

A nuestro parecer, tomó esta decisión para no adentrarse en una cuestión muy peliaguda en aquel período, la versión de elementos religiosos como tales, que convenía evitar para no incurrir en posibles problemas. Esta manera de proceder no representa ninguna novedad, ya que otros traductores de aquellos siglos también se impusieron esta especie de censura. Veamos, a modo de ejemplo, el caso de la traslación al italiano del *Lazarillo de Tormes*.

La novela picaresca *La vida de Lazarillo de Tormes: y de sus fortunas y adversidades*, más conocida como *Lazarillo de Tormes*, vio la luz en 1554. Barezzo Barezzi la trasladó al italiano en 1622, el mismo año en el que se publicó la traducción que estamos estudiando en este trabajo. En italiano se tituló *Il Picariglio castigliano, Cioè la vita di Lazariglio di Tormes nell’Academia Picaresca lo Ingegnoso Sfortunato, Composta, & hora accresciuta dallo stesso Lazariglio, & trasportata dalla Spagnuola nell’Italiana fauella da Barezzo Barezzi* y se imprimió en Venecia “Presso il Barezzi”. Por las razones aducidas con anterioridad, Barezzi decidió cambiar la figura del clérigo de Maqueda por la de un *Medico auaro*, la del fraile mercedario por la de una *Persona qualificata, & a ragion de libri molto dotto* y la del arcipreste de San Salvador por la del *Signor Dottor Ouieda*.

Volvamos a la traducción de Franciosini. Nuestro lexicógrafo, pues, resolvió proponer la dicción *Medico* (1622, pp. 56, 63 y 64, Crusca y Battaglia) cuando hubo de verter no solo la figura religiosa del **Patriarca**¹⁶ (1605, f. 24 v, Covarrubias), sino también la de los **frayles** (1605, f. 28 r, Covarrubias y *Autoridades*) y la de los **religiosos** (1605, f. 28 v, Covarrubias y *Autoridades*).

También prescindió a menudo de los términos religiosos que se ubican en expresiones de juramento (sobre todo) y optó por su supresión directa o por su sustitución por otros carentes de semas relativos a la religión. En

¹⁶ Esta dicción forma parte de la locución adverbial popular **como vn Patriarca** (1605, f. 24 v), según Franciosini *come un Medico* (1622, p. 56). Aquí nuestro traductor no advirtió el sentido subyacente que encierra el significado de **estar**, **ir** o **sentirse** alguien **como vn Patriarca**, en opinión de Rico Manrique (1998), “a sus anchas, muy a gusto”.

nuestro corpus hemos localizado tres de ellas, a saber, **yo te boto a Dios** (1605, f. 37 r; **Dios**, Covarrubias y *Autoridades*), para la que planteó *io ti potrei giurare* (1622, p. 82), **dè al diablo** (1605, f. 36 v; **diablo**, Covarrubias y *Autoridades*), que trasladó como *lasci andar V. S. in malhora* (1622, p. 81; *malora*, Battaglia), y **Yo hago juramento al Criador de todas las cosas, y a los santos quatro Euangelios** (1605, f. 33 r; **Criador** y **Evangelio**, Covarrubias y *Autoridades*), para la que eligió *Io fo giuramento al mio auo, e bisauo, e a quella, che mi legò il bellico* (1622, p. 80; *avo* y *bisavo*, Crusca y Battaglia), es decir, una solución harto distinta a lo que se enunció en el texto original.

6. El léxico de la germanía

A fin de caracterizar a algunos de sus personajes, el Príncipe de los Ingenios empleó en ciertos capítulos de su novela piezas léxicas extraídas de la jerga de la germanía, esto es, un modo de hablar indescifrable creado, a propósito, por los delincuentes y por la gente del hampa para que las demás personas no los comprendieran.

Franciosini se tropezó con dificultades enormes cuando se encaminó a trasladar expresiones pertenecientes a esta esfera. Por ejemplo, le costó trabajo entender las dobles acepciones de muchas de ellas (algo muy complicado de captar incluso en las jergas de hoy en día) y transmitir la vivacidad propia del argot del texto cervantino. Como consecuencia, presentó fragmentos traducidos de manera muy literal carentes, a veces, de sentido y escritos, además, en un italiano estándar que no ayudó a la transmisión de esa vitalidad. Las lexías de este tipo encontradas en nuestro corpus se integran en el campo de la prostitución y del sexo y en el de la delincuencia.

Dentro del primer ámbito, Cervantes denominó a las meretrices, por ejemplo, **mugeres... del partido** o **destraydas**, a juicio de Rico Manrique (1998), “en germanía, prostitutas”. Para la primera estructura, **mugeres... del partido** (1605, f. 5 v, *Autoridades*), nuestro lexicógrafo sugirió *giouanotte... di partito* (1622, p. 11), un concepto incomprensible, y para la segunda, **destraydas** (1605, f. 6 r), se decantó por *vagabonde* (1622, p. 12). Por desgracia, Franciosini no consiguió descifrar el doble sentido de ninguna de esas dos expresiones. Debido a ello, efectuó dos traslaciones literales que provocaron que el lector se sintiera, sobre todo, privado de la sublime ironía que entraña el texto de partida.

Nuestro gramático tampoco detectó las dobles acepciones que escondían el vocablo **solazando** y la estructura **echar vna tela**. Siempre según Rico Manrique (1998), ambas expresiones también aluden a la acción de realizar el acto sexual, significación que se le escapó a Franciosini, quien solo tradujo de modo literal los conceptos más comunes. Así, este escogió

para **solazando** (1605, f. 6 r) *stessero al fresco* (1622, p. 12) y para **echar vna tela**¹⁷ (1605, f. 20 v) *fare vna tela* (1622, p. 70).

Por último, nuestro lexicógrafo se inclinó por verter de forma literal el nombre propio **Esquife** (1605, f. 17 v) como *Eschife* (1622, p. 39), una práctica habitual porque Franciosini adaptó con frecuencia al italiano los topónimos y los antropónimos de la obra original para que estos se pronunciaran de manera similar a los del español. Con respecto a este término, Rico Manrique (1998) defiende que “[...] **esquife** en germanía equivale a ‘rufián’”, un doble sentido cuya traslación comporta no solo para Franciosini, sino incluso para los traductores actuales una dificultad mayúscula, que habría de explicarse mediante la creación de una nota aclaratoria.

Para concluir, hemos de exponer que dentro de la segunda esfera, la de la delincuencia, Franciosini no logró penetrar en las dobles acepciones que contenían dos expresiones, **sanos de Castilla** y **playa de San Lucar**, razón por la cual ofreció, una vez más, opciones traductoras muy literales que causaron una ingente pérdida connotativa y de vivacidad lingüística.

Por lo que atañe a la primera, **sanos de Castilla** (1605, f. 6 v, *Autoridades*), Rico Manrique (1998) elucida que **sano de Castilla** significa en germanía “ladrón disimulado”, tal y como también aparece en *Autoridades*, aspecto que nuestro gramático no percibió porque sin añadir nota marginal alguna, como en cambio sí hizo en otras ocasiones, trasladó esa pieza léxica como *sempliciotti di Castiglia* (1622, p. 14; *sempliciotto*, Crusca y Battaglia). Y por lo que concierne a la segunda, nos hemos percatado de que Franciosini actuó de igual modo. En efecto, nuestro lexicógrafo eligió para traducir **playa de San Lucar** (1605, f. 6 v), según Rico Manrique (1998), “en tiempos de Cervantes, punto de reunión de pícaros, indeseables y fugitivos de la justicia”, un simple *riuiera di San Lucar* (1622, p. 14), sin que imaginara que muchas expresiones cervantinas encubren más de un significado, juegos de palabras, etc., lo cual nos induce a pensar que, como sostienen algunos críticos, sus conocimientos sobre nuestro idioma no siempre eran sublimes.

7. Conclusiones

En algunas ocasiones, Franciosini explicitó las dificultades que comporta la actividad traductora. Por ejemplo, así se expresó al principio de sus *Dialogos apazibles* al dirigirse “Al Signor Lettore” (1626, p. 6):

¹⁷ En opinión de Rico Manrique (1998), “**tela**, en germanía, es también ‘coito’”.

Prima che tu tocchi (Signor Lettore) il polso a questi miei Dialoghi, ti voglio dire la loro infermità. Sono tradotti, e per conseguenza corrotti; e non stò punto in dubbio, che nello Spedale del tuo giudizio, non habbiano à parere Stroppiati.

Y de esta manera justificó en su traslación del *Quijote*, en concreto en la advertencia “A’ cvriosi lettori”, que optó por no verter los versos (Franciosini, 1622, f. a 3 r):

I versi non gl’hò tradotti, perche oltre all’esser difficile à chi non è Poeta; non mi son parsi tanto essenziali alla dichiarazion della prosa, che questa non si sia senz’essi, potuta volgarizzare.

Pues bien, tras haber examinado en profundidad los diez primeros capítulos de la primera parte de su versión del *Quijote* cervantino (aunque también hemos leído la traslación completa), hemos de afirmar que no le faltaba razón. En efecto, nuestro gramático hubo de hacer frente a numerosos problemas que, por desgracia, no siempre supo solucionar, ya que, por ejemplo, propuso a menudo traducciones excesivamente literales. Además, tampoco logró transmitir siempre el plurilingüismo de Cervantes.¹⁸ Pero vayamos por partes.

Franciosini realizó con cierta frecuencia propuestas traductoras harto literales sobre determinadas expresiones del texto original, algo que ya advirtió en su portada de 1622, cuando comentó que la novela la trasladó “con fedeltà” (a nuestro parecer, “literalmente”). Los principales obstáculos los encontró cuando debió verter las unidades fraseológicas (en especial los proverbios) que emplea Sancho en numerosos pasajes, los juegos de palabras, los elementos culturales, los dobles sentidos, las deformaciones lingüísticas (sobre todo sanchopancescas) y la terminología proveniente de las jergas. Como consecuencia de sus decisiones, hallamos una obra que, en numerosas partes (máxime en los últimos capítulos de la segunda parte), suele carecer de la expresividad, la vivacidad y la frescura cervantinas. No obstante, hemos detectado que para intentar mantenerlas y, también, para no perder el humor y la ironía de la novela del Príncipe de los Ingenios, Franciosini se sirvió a veces del léxico toscano, ya que, al igual que sucede con otros dialectos presentes en Italia, encerraba correspondencias mucho más idóneas en relación al texto original que el italiano estándar.

¿Y qué ocurre con la traslación del plurilingüismo que ostenta nuestro genio alcalaíno? Por lo que concierne al lenguaje que utilizan los dos personajes principales, Don Quijote y Sancho, Franciosini no erró casi nunca cuando se ocupó de verter la lengua culta, arcaica y afectada del primero. En

¹⁸ Y eso sin contar con que a veces ni siquiera comprendió determinados fragmentos de la novela cervantina, lo cual conllevó la presencia de errores de calado, ni con que incluso se inventó lexías ausentes en los diccionarios italiano de la época.

cambio, se topó con más escollos cuando afrontó la traducción del lenguaje espontáneo, rural y coloquial de Sancho y, además, de todo lo cotidiano que contiene la obra. El resultante es la utilización, por parte de Sancho, de un italiano muy estándar, incapaz de reflejar toda la comicidad, la vivacidad, la expresividad y todo el humorismo inherentes a la novela cervantina. También aquí, para intentar compensar estas carencias, debidas en parte a la naturaleza del idioma italiano, asunto en el que no podemos adentrarnos en estas páginas, nuestro lexicógrafo recurrió en ocasiones al uso del léxico toscano, ya que, y volvemos a repetir lo que hemos comentado pocas líneas más arriba, en los dialectos existentes en Italia se hallaban (y se hallan) no pocas expresiones que reflejan mucho mejor que el italiano estándar esa cotidianeidad, esa naturalidad.

Nota biográfica: José Francisco Medina Montero, licenciado, licenciado de Grado y doctor (Filología Hispánica, opción Lengua Española) por la Universidad de Extremadura, es Catedrático de Lengua, Traducción y Lingüística española del Dipartimento di Scienze Giuridiche, del Linguaggio, dell'Interpretazione e della Traduzione de la Universidad de Trieste (Italia). Sus líneas de investigación son la Lengua Española (diacronía y sincronía) como primera y segunda lengua, la Lingüística contrastiva italiano-español-italiano y la Traducción italiano-español-italiano. En estos campos ha publicado más de 60 trabajos entre artículos, libros, etc. Es miembro y/o coordinador de 18 proyectos de investigación financiados, pertenece a 6 comités editoriales y científicos de revistas, ha formado parte de 33 comités organizadores y científicos de congresos internacionales y ha sido ponente de más de 80 conferencias en Europa y Latinoamérica. En la Universidad de Trieste ha ocupado los cargos de Director de Grado, de Vicerrector de Movilidad Internacional y de Relaciones Internacionales y de Director del Centro de Idiomas.

Correo electrónico: jmedina@units.it

Bibliografía

- Anónimo 1554, *La vida de Lazarillo de Tormes: y de sus fortunas y adversidades*, Juan de Junta, Burgos. Traducción de Barezzi B. 1622 titulada *Il Picariglio castigliano, Cioè la vita di Lazariglio di Tormes nell'Academia Picaresca lo Ingegnoso Sfortunato, Composta, & hora accresciuta dallo stesso Lazariglio, & trasportata dalla Spagnuola nell'Italiana fauella da Barezzo Barezzi*, Barezzi, Venetia, https://www.google.it/books/edition/Il_Picariglio_castigliano_cioe_La_vita_d/H8pGpNwq2hUC?hl=it&gbpv=1 (consultado en febrero de 2024)
- Battaglia S. 1964, *Grande Dizionario della Lingua Italiana*, UTET, Torino.
- Bernardi D. 1995, *El “Don Chisciotte” de Lorenzo Franciosini (1622): un caso di (auto)censura*, en Romero Muñoz C., Pini Moro D. y Cancellier A. (eds.), *Atti delle Giornate cervantine: Venezia, 7 maggio 1991 (II Giornata); Padova, 4 maggio 1992 (III Giornata); Venezia, 23 aprile 1993 (IV Giornata)*, Unipress, Padova, pp. 93-104.
- Botta P. 2019, *La traducción del “Quijote” al italiano*, en Botta P. (coordinadora), *Las traducciones del “Quijote”: homenaje al hispanismo internacional*, Fundación Duques de Soria y Ediciones Universidad de Valladolid, Valladolid, pp. 87-96.
- Cervantes Saavedra M. de 1605, *El ingenioso hidalgo Don Qvixote de la Mancha*, Iuan de la Cuesta, Madrid, <http://bdh-rd.bne.es/viewer.vm?pid=d-1804836> (consultado de julio a diciembre de 2023). También hemos recurrido con frecuencia a la edición del Instituto Cervantes dirigida por Rico Manrique F. 1998, Instituto Cervantes, Madrid, <http://cvc.cervantes.es/literatura/clasicos/quijote/> (consultado de julio a diciembre de 2023).
- Cervantes Saavedra M. de 1605, *El ingenioso hidalgo Don Qvixote de la Mancha*, Iuan de la Cuesta, Madrid. Traducción de Franciosini L. 1622 titulada *L'ingegnoso Cittadino Don Chisciotte della Mancia*, Andrea Baba, Venetia, <http://books.google.es/books?id=9wpEAAAACAAJ&hl=ca&pg=PA1#v=onepage&q&f=false> (consultado de julio a diciembre de 2023).
- Colón i Domènech G. 1974, *Die ersten romanischen und germanischen Ueberstzungen des “Don Quijote”*, Francke, Bern.
- Covarrubias y Orozco S. de 1611, *Tesoro de la lengua castellana, o española*, Luis Sánchez, Madrid, <http://books.google.es/books?id=K10MJdL7pGIC&printsec=frontcover&dq=covarrubias+1611&hl=es&sa=X&ei=M8dxVNDaCYP8ygPdQoKwBQ&ved=0CCIQ6AEwAA#v=onepage&q=covarrubias%201611&f=false> (consultado de julio a diciembre de 2023).
- Da Montemerlo G.S. 1594, *Thesoro della lingua toscana*, Giacomo Antonio Somascho, Venetia, https://books.google.it/books?id=ns-i07Aa_H4C&pg=PA899&dq=%22con+cento+buon+anni%22&hl=it&sa=X&redir_esc=y#v=onepage&q=%22con%20cento%20buon%20anni%22&f=false (consultado en febrero de 2024).
- Del Bravo S. 2024, *Lorenzo Franciosini*, en *Dizionario Biografico Treccani*, Istituto della Enciclopedia Italiana, Roma, [https://www.treccani.it/enciclopedia/lorenzo-franciosini_\(Dizionario-Biografico\)/?search=FRANCIOSINI%2C%20Lorenzo](https://www.treccani.it/enciclopedia/lorenzo-franciosini_(Dizionario-Biografico)/?search=FRANCIOSINI%2C%20Lorenzo) (consultado en enero de 2024).
- Elizalde Armendáriz I. 1981, “*El Quijote*” y la novela moderna, en Criado de Val M. (ed.), *Cervantes, su obra y su mundo: actas del I Congreso Internacional sobre Cervantes*, Edi-6, Madrid, pp. 949-962.

- Fernández S.J.J. 2008, *Bibliografía del “Quijote” por unidades narrativas y materiales de la novela*, Centro de Estudios Cervantinos, Madrid.
- Franciosini L. 1620, *Vocabolario italiano, e spagnolo*, Angelo Ruffinelli & Angelo Manni, Roma, https://books.google.it/books?id=bBAScLv0F54C&printsec=frontcover&hl=it&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false (consultado en enero de 2024).
- Franciosini L. 1624, *Gramatica spagnola, e italiana*, Giacomo Sarzina, Venetia, https://books.google.it/books?id=ykdXwAEACAAJ&printsec=frontcover&hl=it&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false (consultado en enero de 2024).
- Franciosini L. 1626, *Dialogos apazibles, compvestos en Castellano, y traduzidos en Toscano*, Giacomo Sarzina, Venetia, https://www.google.it/books/edition/Dialoghi_piacevoli/RN9DAAAACAAJ?hl=it&gbpv=1&dq=inauthor:%22Lorenzo+Franciosini%22&printsec=frontcover (consultado en enero de 2024).
- Garatea Grau C. 2007, *En torno a la lengua de don Quijote y Sancho*, en Lara Ramos L.F., Ortega Ortiz R.Y. y Tenorio Trillo M.L. (eds.), *De amicitia et doctrina. Homenaje a Martha Elena Venier*, El Colegio de México, México, pp. 171-186.
- Luttikhuisen F. 2008, *Breve aproximación a la primera traducción neerlandesa del “Quijote”*, en Pedraza Jiménez F.B. y González Cañal R. (eds.), *Con los pies en la tierra: Don Quijote en su marco geográfico e histórico. Homenaje a José M^a Casasayas*, Universidad de Castilla-La Mancha, Cuenca, pp. 305-316.
- Martínez Egido J.J. 2010, *La obra pedagógica del hispanista Lorenzo Franciosini (un maestro de español en el siglo XVII)*, Polimetrica, Monza-Milano.
- Medina Montero J.F. 2015, *El primer capítulo de la primera parte del “Quijote” de Franciosini: observaciones sobre la traducción de algunos elementos*, en “Rassegna Iberistica” 38, 104, pp. 203-224.
- Medina Montero J.F. 2016, *El segundo capítulo de la primera parte del “Quijote” de Franciosini: observaciones sobre la traducción de algunos elementos*, en “Rivista Internazionale di Tecnica della Traduzione” 18, pp. 147-159.
- Medina Montero J.F. 2017, *El tercer y el cuarto capítulo de la primera parte del “Quijote” de Franciosini: observaciones sobre la traducción de algunos elementos*, en Luque Toro L. y Luque R. (eds.), *Léxico español actual V*, Cafoscarina, Venezia, pp. 171-186.
- Medina Montero J.F. 2018a, *El quinto y el sexto capítulo de la primera parte del “Quijote” de Franciosini: observaciones sobre la traducción de algunos elementos*, en “Artifara” 18, monográfico *Traducción e intertextualidad; aspectos sincrónicos y diacrónicos*, pp. 9-22.
- Medina Montero J.F. 2018b, *El séptimo y el octavo capítulo de la primera parte del “Quijote” de Franciosini: observaciones sobre la traducción de algunos elementos*, en “Artifara” 18, monográfico *Traducción e intertextualidad; aspectos sincrónicos y diacrónicos*, pp. 25-39.
- Medina Montero J.F. 2019, *El noveno y el décimo capítulo de la primera parte del “Quijote” de Franciosini: observaciones sobre la traducción de algunos elementos*, en “Rivista Internazionale di Tecnica della Traduzione” 21, pp. 61-78.
- Pini Moro D. y Moro G. 1992, *Cervantes in Italia: contributo a un saggio bibliografico sul cervantismo italiano (con un'appendice sulle trasposizioni musicali)*, en Pini Moro D. (ed.), *Don Chisciotte a Padova (Atti della I Giornata Cervantina, Padova, 2 Maggio 1990)*, Editoriale Programma, Padova, pp. 149-268.

- Real Academia Española 1726-1739, *Diccionario de Autoridades*, Real Academia Española, Madrid, <http://web.frl.es/DA.html> (consultado de julio a diciembre de 2023).
- Repetti E. 1845, *Supplemento al Dizionario geografico fisico storico della Toscana*, Giovanni Mazzoni, Firenze,
https://books.google.it/books?id=dXZJPftWjykC&pg=PA59&lpg=PA59&dq=%22fra+puccino%22+firenze&source=bl&ots=adVbf_8muW&sig=FtPSiEbQIU9kcjZh43jtvngzz4&hl=it&sa=X&ved=0ahUKEwjnnI2erbDRAhXCShQKHYINDp4Q6AEIHZAB#v=onepage&q=%22fra%20puccino%22%20firenze&f=false (consultado en enero de 2024).
- Riquer y Morera M. de. 1942, *La obra del hispanista Lorenzo Franciosini, primer traductor del “Don Quijote” al italiano*, en “Revista Nacional de Educación” 21, pp. 21-28.
- San Vicente Santiago F. 2016, *Lorenzo Franciosini, “Gramatica spagnola e italiana”, Venezia, Sarzina, 1620. Introducción y criterios de transcripción*, en *La tradición gramatical del español en Italia. La “Gramatica spagnola e italiana” de Lorenzo Franciosini. Estudio y edición crítica*, Cleup, Padova, pp. 1-10, <http://www.epigrama.eu> (20.01.2024).
- Scaramuzza Vidoni M. 2013, *Il primo “Don Chisciotte” bilingue in Italia*, en “Tintas. Quaderni di letterature iberiche e iberoamericane” 3, pp. 221-227.
- Vega Cernuda M.Á. 2005, *La traducción del “Quijote” o ¿qué “Quijote” leen los europeos?*, en Vega Cernuda M.Á. (ed.), *¿Qué “Quijote” leen los europeos?*, Instituto Universitario de Lenguas Modernas y Traducción, Madrid, pp. 1-35.
- VVAA 1612, *Vocabolario degli Accademici della Crusca*, Venetia, http://www.lessicografia.it/ricerca_libera.jsp (consultado de julio a diciembre de 2023).

DISCURSO MUSEÍSTICO E IDENTIDAD

Una comparación

SIMONE TEPEDINO
UNIVERSITÀ DI MODENA E REGGIO EMILIA

Abstract – The purpose of this work is to compare the websites of two main Spanish museums, El Prado and Reina Sofia, with the aim of identifying similarities and differences in their discursive strategies. After defining the criteria of the corpus selection and its analysis (qualitative, inductive and synchronic), and specifying the definition of the discourse related to the auto-representation of the museums, the comparison will be carried out. Its goal is to demonstrate that both museums use their sites to promote their own personalities mainly through three sections, *Historia*, *Misión* and *Educación*, in a bidirectional and mutual relation with their community of visitors. Moreover, this study can further demonstrate that multimodal texts, such as websites, undoubtedly feature a social practice which is considered as the milestone of discourse analysis.

Keywords: museum auto-representation; discourse analysis; multimodal texts; discursive strategy; communication.

1. Introducción: el discurso digital en el ámbito museístico

En tiempos recientes el desarrollo, cada vez más extenso, de la comunicación en la web ha engendrado un correspondiente desarrollo del discurso digital. La importancia de dicho discurso estriba en el cambio fundamental que ocurrió en la manera de interactuar de los seres humanos. La comunicación ya no es directa como en el pasado: es lo que Susan C. Herring define Computer-mediated communication (CMC), “a communication that takes place between human beings via the instrumentality of computers” (Herring 1996, p. 1).

Por lo tanto, indagar sobre el discurso digital significaría analizar la comunicación entre individuos a través de medios como chats, sistemas de mensajería instantánea o blogs, que imitan la comunicación oral al tener los rasgos típicos de una conversación¹. En otras palabras, existe una “co-

¹ A este propósito, Halliday ofreció una distinción entre lengua oral y lengua escrita, considerándolas como dos facetas fundamentales de la comunicación humana. La lengua escrita, aun no teniendo los elementos prosódicos y paralingüísticos de la oralidad, se caracteriza por una

presencia temporal” (Mancera Rueda, Pano Alamán 2013, p. 11), una interacción directa y sincrónica entre los participantes en una conversación virtual.

Sin embargo, es posible encontrar otras formas de discurso presentes en la web, como por ejemplo los sitios web. En este caso la comunicación puede definirse como indirecta y asíncrona: no existe una inmediatez en los intercambios de los internautas. A veces, simplemente no hay una comunicación al no existir un sistema de mensajería (que, en cambio, se puede detectar de manera asíncrona en los foros o los correos). Empero, eso no significa que no tenga lugar una comunicación en un sitio web: por el contrario, existe y se puede resumir tomando el esquema de comunicación de Roman Jakobson.



Imagen 1
El esquema comunicativo de Roman Jakobson.

mayor densidad léxica; es decir, las palabras poseen un amplio contenido dentro de un conjunto abierto, como en el caso de la palabra ‘puerta’, que puede incluirse en un conjunto casi ilimitado con términos como ‘portón’, ‘ventana’ etc. (Halliday 1989). En el caso del presente trabajo, como se podrá comprobar en las próximas páginas, por cierto muchas de las partes escritas tienen esta densidad; sin embargo, la multimodalidad del texto complica su análisis. Además, en tiempos más recientes, Mancera Ruda y Pano Alamán (2013) han examinado las formas del español coloquial en la red, distinguiendo entre CMO (Comunicación Mediada por el Ordenador) y ADMO (Análisis del Discurso Mediado por el Ordenador). La primera representa un conjunto de modalidades de interacción surgidas con las nuevas tecnologías; la segunda es un método de análisis de las propiedades pragmáticas y lingüísticas de la CMO. Las autoras concluyen sosteniendo que hay cierta correspondencia entre el sistema oral y el escrito en las redes sociales: en efecto, el segundo busca las características del primero a través de los signos gráficos utilizados (que imitan los fonemas) de la informalidad y de la inmediatez de los mensajes. En última instancia – y sin alguna pretensión de exhaustividad sobre el tema –, Kress subraya cómo, en la web, el texto escrito puede muchas veces ser reemplazado por la inmediatez de una imagen (Kress 2003, pp. 32-33).

Este esquema² ofrece una síntesis precisa de los elementos que concurren en la formación de cualquier comunicación humana y, por consiguiente, introduce la idea de que existe una dimensión pragmática también en las páginas web. Los componentes del modelo son:

- Emisor: quien elabora el mensaje. En el caso del presente estudio, quien administra la página web.
- Receptor: los internautas.
- Mensaje: lo que se escribe en las páginas que se analizarán más adelante.
- Contexto: el momento en el cual se produce y/o se lee el mensaje.
- Código: multimodal (texto e imágenes, quizá también videos).
- Canal: la pantalla.

En resumen, la diferencia con la comunicación oral estriba en el hecho de que no hay una respuesta inmediata del receptor. Esto lleva a interrogarse sobre: ¿cómo se pueden conocer las respuestas? O, en otras palabras, ¿cómo se puede evaluar la eficacia de una página web en su comunicación con los internautas / la comunidad de referencia? ¿Cuál sería su representación frente a dicha comunidad? Además, ¿qué relación hay entre página web y comunidad de referencia, y cómo se construye?

A estas preguntas se intenta contestar en el presente trabajo. Dentro del ámbito museístico, se analizará un corpus de páginas web del Museo del Prado y del Museo Reina Sofía³ que se compararán, con el fin de comprobar, por un lado, la representación que cada institución ofrece de sí misma, y, por otro las diferencias dentro de esta representación.

Como conclusión de esta parte introductoria, cabe explicar que esta publicación surge del trabajo del grupo de investigación de la Universidad de Modena y Reggio Emilia llamado CAP (Comunicazione Accademica e Professionale). Formé parte de este grupo como becario colaborador en 2021. Su objetivo es analizar la comunicación en varios ámbitos (y en varios idiomas: inglés, español, francés y alemán) entre los cuales destacan:

[...] la comunicazione aziendale e istituzionale (istituzioni politiche e culturali): a) comunicazione della corporate social responsibility e il suo ruolo nella (ri-)costruzione della fiducia nella comunicazione aziendale; b) analisi di siti istituzionali nazionali e transnazionali; c) analisi comparativa di siti e materiali nella comunicazione web museale, anche rivolti all'infanzia e all'adolescenza. (Del web de Unimore⁴)

² El esquema, extraído del sitio <http://creacionverbal.blogspot.com/2015/03/el-esquema-de-la-comunicacion-de-roman.html> (08.04.2023), resume bien la teoría que Jakobson presenta en su artículo *Lingüística y poética* (Jakobson 1985).

³ En las próximas páginas se explicará mejor el tamaño del corpus y la razón de su elección.

⁴ <https://www.dslc.unimore.it/site/home/ricerca/progetti-in-corso.html> (06.04.2023)

El grupo de español se centró en la comunicación en ámbito museístico. El objetivo final del CAP es la creación de una base de datos, un corpus en los distintos idiomas útil a los futuros desarrollos del Departamento de Estudios Lingüísticos y Culturales de la Universidad de Modena y Reggio Emilia.

1.1. Características del discurso digital

Como se ha afirmado en el apartado precedente, con el desarrollo de la web se ha creado un nuevo tipo de discurso, el discurso digital. Antes de proceder con nuestro análisis, es necesario establecer las características de este discurso para que sea claro el objeto de estudio.

Ante todo, el rasgo fundamental es la intertextualidad, es decir, la posibilidad de acceder directamente a otras secciones del sitio web (u otros sitios) a través de los enlaces – y acceder, pues, a la información – con una rapidez notable⁵. Por lo tanto, el receptor es un internauta que se mueve en un espacio mediado por la pantalla, es decir, desterritorializado.

[...] navegar a través de un hipertexto es moverse en un mundo de libertad en donde el sujeto es, al mismo tiempo, lector y creador; este *lectoescritor* es una nueva figura híbrida, heterogénea, casi mestiza, producto especialmente de la *interactividad* [...], *conectividad* [...], *multimedialidad* [...] y de la *multisecuencialidad* [...] (Covandonga 2014, p. 179. Cursiva presente en el original).

Existen, sin duda, otras características que se consideran fundamentales al tratar del discurso digital:

1. Multimodalidad.
2. Continua transformación.
3. Adaptación a los cambios/valores sociales.

El primer punto (multimodalidad) tiene que ver con la naturaleza múltiple del código del sitio internet⁶. Siguiendo el enfoque de Teun Van Dijk, se puede afirmar que el discurso del web:

⁵ El concepto de intertextualidad, desde un punto de vista más cognitivo, puede traducirse en lo que Lepkowska-White e Imboden definen ‘mystery’ (2013, p. 291), es decir, la capacidad que el sitio tiene de fomentar la curiosidad del usuario para navegar a través de los enlaces. Este concepto se une a otros tres (coherency, complexity, legibility) que representan los elementos básicos para que un sitio web sea construido de manera eficaz (Lepkowska-White, Imboden 2014).

⁶ Según Carliner, Shi, y Wan, la multimodalidad es una de la tres affordances que representan un texto web, junto con hypertextuality y functionality. La primera es la capacidad del texto de tener varios códigos semióticos; la segunda define el grado de interconectividad de las varias partes del sitio o de los mensajes dentro de una cadena de comunicación; y, finalmente, la tercera se refiere a las funciones tecnológica del medio que se utiliza (browsing, searching, interacting, collaborating). (Carliner, Shi, Wan, 2020, pp. 281-282).

No es solamente oral y verbal; como texto escrito, también está caracterizado por importantes variaciones de tipografía (como en las negritas [...]), imágenes [...], música y otros sonidos, así como otros tipos de signos «encarnados» (*embodied*), como gestos, expresiones faciales, posición del cuerpo, etc., observables en la interacción hablada y que la **semiótica** del discurso estudia (Van Dijk 2016, p. 21. Cursiva y negrita presente en el original).

Todo eso supone una dificultad más en el estudio multimodal del discurso, como existen distintos canales semióticos que lo componen. Por ende, es necesario considerar el texto en su complejidad, en su carácter multimodal, para entender plenamente el significado que el emisor quiere transmitir. El análisis, entonces, tiene múltiples facetas que no se pueden excluir, que Kress (2003, p. 19) resume al tratar de la nueva lógica de la imagen en el estudio de la página web.

The logic of writing is temporal/sequential; the elements of speech certainly, and those of writing in a quasi-fashion, unfold in time and in sequence. [...] The logic of image on the other hand is spatial/simultaneous. All the elements of the image are related in spatial arrangement, and they are simultaneously present (Kress 2003, p.19).

Eso significa que una página web, por su naturaleza, organiza su disposición según la lógica de la imagen: varios códigos semióticos están presentes al mismo tiempo según un orden espacial, desvelándose simultáneamente. Es un nuevo paradigma que hace que el análisis sea más complejo.

El segundo punto está representado por la capacidad de la web de transformarse, de adaptarse rápidamente a los cambios de la sociedad. Dicha capacidad se refleja en las frecuentes variaciones que experimentan las páginas web, aun en poco tiempo: es la continua transformación de las páginas web. En efecto, es muy común para el internauta darse cuenta de que un sitio web ha cambiado algún aspecto: textos, colores, estructuras. Esta es también la razón por la cual, por ejemplo, en los artículos académicos se pone la fecha de consulta de un determinado sitio: la página estudiada podría haber sufrido modificaciones en el momento de la lectura del texto. En primer lugar, esta característica representa un problema serio para el investigador, dado que su análisis podría desmontarse por razones independientes de su voluntad. Este punto fue uno de los más debatidos en las reuniones de los miembros del CAP: sería útil, para el investigador, tener un recurso informático que permita guardar las páginas, manteniéndolas como en el momento en el que se consultaron. Un programa de este tipo podría también facilitar un estudio de tipo histórico-cronológico del web, dado que sería posible archivar todas las secciones analizadas, siguiendo posibles cambios

de texto, estructura, colores etc. El grupo de investigación ha identificado en el programa ArchiveBox una posible solución:

ArchiveBox is a powerful, self-hosted internet archiving solution to collect, save, and view sites you want to preserve offline. [...] The goal is to sleep soundly knowing the part of the internet you care about will be automatically preserved in durable, easily accessible formats for decades after it goes down (del sitio de ArchiveBox⁷).

Esta tecnología, todavía por profundizar y poner en marcha adecuadamente por parte del grupo de investigación, podría representar un importante desarrollo futuro para quien decide adentrarse en la investigación lingüística y semiótica de las páginas web. Sin embargo, no estando plenamente perfeccionada, en el presente trabajo se ha decidido consultar las páginas web y guardarlas sin la ayuda de ArchiveBox.

En segundo lugar, y desde un punto de vista práctico, la continua transformación de las páginas web acerca el discurso digital al oral. De hecho, Volli habla de una:

[...] perenne trasformabilità dei contenuti testuali – che non hanno più la definitiva stabilità cui il procedimento di stampa li innalzava, o li condannava. E queste sono tutte caratteristiche tipiche della comunicazione orale, e di forme di vita sociale che fino a poco tempo fa credevamo appartenere al nostro passato piuttosto che al nostro futuro (Volli 2003, p. 197).

La paradoja, según Volli, estriba en la inestabilidad del discurso digital que lo acercaría a la oralidad: un tipo de comunicación que, en cambio, parecía destinada a reducirse mucho con la web. La presencia de los enlaces y los continuos cambios de las páginas web conllevan la posibilidad de modificar ininterrumpidamente un texto escrito, algo imposible con un texto impreso. En este sentido, dicho texto sería volátil, precisamente como la comunicación oral o, en palabras de Gianmatteo, Gubitosi y Parini (2017, p.14) un texto dinámico, “cuya esencia implica una constante actualización”.

Esta particular comunión entre discurso digital y oral refuerza la idea de la dimensión pragmática del presente estudio y, a su vez, conduce al último punto tratado en esta introducción, es decir, la adaptación a los cambios y valores sociales del discurso digital.

La posibilidad de transformar una página web por varias razones (elección del emisor del mensaje, cambios de estrategia o, en el caso del museo, presentación de una nueva muestra, por ejemplo) permite a las páginas web también seguir, o quizás adelantar, los cambios sociales a través de internet.

⁷ <https://archivebox.io/> (22.04.23).

En un ininterrumpido diálogo con el internauta, que se mueve dentro de este espacio desterritorializado, el discurso digital se modifica, cambia, de acuerdo también con lo que ocurre en un determinado momento. Esta sería la razón principal por la cual el enfoque pragmático funciona para este estudio: precisamente porque permite analizar mejor el discurso digital como expresión de aquella particular comunicación que ocurre entre el museo y su comunidad de referencia. La cual representaría, en términos más cercanos a la literatura, el lector ideal o idealizado: es decir, un constructo (Rainer 1979, p. 124) al cual el emisor se dirige.

Dentro de este marco, como ya se ha dicho, con respecto al español se decidió focalizar la atención del estudio en la comunicación en el ámbito museístico. Más específicamente, se recogió un corpus de los más importantes museos españoles con el objetivo de averiguar las distintas estrategias a través de las cuales las instituciones promueven a sí mismas en el web⁸.

Ya a partir de esta última frase se pueden entender los problemas que una indagación como esta puede plantear:

1. Definir la importancia de los museos: esto es, el criterio de elección del corpus.
2. El método de análisis del corpus mismo.

Estos serán precisamente los puntos que se tratarán a continuación.

2. Definición del corpus y metodología de análisis

Al momento de elegir qué instituciones incluir en el corpus de los museos a analizar, decidimos basarnos en un artículo de “El País” que trataba precisamente el tema de la popularidad de los museos en España. Su título es *Los museos españoles, imparables en 2019: en torno a 20 millones de visitantes* (“El País” 2020⁹).

⁸ Analizar páginas web significa adentrarse en un género específico. Género, en este caso, es una palabra que puede definirse, parafraseando a Calvi (2016), como una clase de textos con un propósito comunicativo, creada socialmente y clasificable a partir de criterios intratextuales, extratextuales y funcionales. De este modo se plasma un sistema de conocimientos. En nuestro caso, todo eso significa que los sitios museísticos quieren comunicar las características de una determinada institución a su comunidad de referencia con un determinado registro, que proviene tanto del ámbito del arte como del ámbito publicitario, empleando al mismo tiempo herramientas del discurso histórico, al tratar de la evolución del museo. Además, no se olvide la multimodalidad del discurso, concepto que ya se ha tratado y que volverá en las próximas páginas.

⁹ Extraído del sitio web de “El País”, https://elpais.com/cultura/2020/01/03/actualidad/1578045917_396367.html. (07/04/2023)

En este escrito, publicado en 2020, se ofrecía una clasificación de los museos más visitados en el año anterior. Entre ellos, había: El Prado, Reina Sofía, Caixa Forum y Thyssen (Madrid); Guggenheim (Bilbao); MNAC y MACBA (Barcelona); y, finalmente, IVAM (Valencia).

El artículo de “El País” permite tener un grupo de museos que, a través de un criterio específico – una clasificación de popularidad – representan los más importantes de la península ibérica. Este concepto puede adaptarse a lo que Rojo define como tamaño, representatividad y equilibrio del corpus, es decir, la capacidad del corpus de “reflejar las características de la población de la cual ha sido extraída en los parámetros pertinentes” (2021, p. 291). Aquí es necesaria una digresión para explicar mejor el objetivo del presente trabajo: nuestro análisis que, como se explicará más adelante, es cualitativo, quiere desvelar las estrategias a través de las cuales los museos se auto-representan frente a su comunidad de referencia, con la que tienen una relación biunívoca. En este caso, la representatividad del corpus¹⁰ puede, entonces, coincidir con la popularidad de los museos: aquella respuesta al sitio internet que se mencionaba en la introducción de este trabajo sería, pues, la presencia del público. Eso demostraría la eficacia de la comunicación de una determinada institución unida, evidentemente, a su colección.

Una vez establecido el criterio de elección del corpus cabe definirlo. En palabras de Winnie Cheng sería “[...] a collection of texts that has been compiled for a particular reason [...]. A corpus, then, is a collection of texts based on a set of design criteria, one of which is that the corpus aims to be representative” (Cheng 2012, p. 3)¹¹.

La capacidad del corpus de ser representativo y sus características plantean muchas dudas en el investigador. Como especifica Cheng (2012, p. 4), hay que considerar:

1. Si los textos tienen que ser elaborados de manera natural (‘naturally occurred’), es decir, dentro de un contexto de producción que sea espontáneo, no construido.
2. Si es necesario incluir textos completos, o se pueden también incluir partes de ellos.
3. El contexto en el cual se produce el texto: cuándo ocurrió (y de qué tipo es), quiénes son los participantes y su objetivo comunicativo.

¹⁰ El corpus recogido consta de 26 páginas web y un total de 4819 palabras.

¹¹ Esta definición de corpus, bastante general, podría acompañarse por la de Guillermo Rojo: “un conjunto de (fragmentos de) textos, orales o escritos, producidos en condiciones naturales, conjuntamente representativos de una lengua o de una variedad lingüística, en su totalidad o en alguno(s) de sus componentes, que se almacenan en formato electrónico y se codifican con la intención de que puedan ser analizados científicamente (2021, p. 1). Aun siendo más detallada, se ha decidido adoptar la de Cheng porque, por su naturaleza general, mejor se adapta a los objetivos del presente trabajo.

4. La categoría de las palabras analizadas, es decir, sus accidentes gramaticales. Las dudas estriban, en este caso, en la manera en la cual anotar variaciones en el corpus. Evidentemente, este último punto es útil solo para análisis de tipo cuantitativo.

Cada uno de estos puntos – excepto el 4, y se explicará la razón – puede ser analizado en relación con la recogida del corpus del presente trabajo.

Ante todo, los sitios web de los museos se construyen dentro de un contexto natural. No son espontáneos como los textos de una conversación, dada la naturaleza escrita del sitio internet: sin embargo, se producen con una “intención real de comunicar algo” (Rojo 2021, p., 1), es decir, promover el museo dirigiéndose a un público objetivo muy amplio (véanse otras instituciones, como las escuelas, pero, en general, se podría incluir toda la comunidad que rodea al museo).

A este propósito, cabe añadir algunas consideraciones de carácter no exclusivamente lingüístico que ayudarán a comprender mejor la estrategia que subyace a la estructura y a la planificación de un sitio museístico. Su objetivo principal es ofrecer al usuario “emotional appeal, aesthetic experience, or a good impression” (Yu 2021, p. 251). Es decir, la parte emocional es la que predomina al construir un sitio de un museo, que puede considerarse un promotor de una peculiar experiencia para el usuario. Dicha experiencia se divide en cuatro partes: entretenimiento, educación, evasión y estética (Pallud, Straub 2014, p. 360), elementos que concurren en la formación del sitio y fortalecen su relación con la comunidad de referencia del museo – dentro del contexto natural mencionado antes. Además, como subrayan Pallud y Straub (2014, p. 361), este contexto incluye un cambio de paradigma que ha ocurrido en los últimos años: el museo ha pasado a ser de ‘collection-driven’ a ‘visitor-centered’, evidenciando todavía más como la relación entre emisor y receptor del mensaje desempeña un papel de absoluta importancia. Como conclusión de esta breve digresión, se puede afirmar que al planificar un sitio, su gráfica, su estructura, es fundamental centrarse en la comunidad de referencia para tener un impacto en su parte más emocional, con el fin de construir un sitio eficaz y representativo de la identidad de un determinado museo.

En segundo lugar, en nuestra investigación se decidió considerar las páginas web completas, con todos sus párrafos e imágenes, desarrollando pues un análisis de tipo multimodal: como afirma López Pena, “el enfoque multimodal se preocupa del significado que se crea a través de las diferentes combinaciones e interacciones que se pueden tener, en un mismo mensaje” (López Pena 2021, p. 840). Eso conlleva dos observaciones: primero, el análisis debería desarrollarse de modo cualitativo y no cuantitativo. Incluir textos completos ayuda a tener una visión amplia y general, a comprender mejor las estrategias de promoción de la “personalidad” del museo, que tiene una relación bidireccional con su comunidad de referencia. Un análisis

cuantitativo no parece responder adecuadamente al objetivo del presente trabajo, al ser normalmente un instrumento para “utilizzare statisticamente informazioni raccolte attraverso testi”, o cuando “serve estrarre dati da corpora testuali” (Cortellazzo 2013, p. 300). La oportunidad que se plantea es la de ir “beyond the sentence” (Biber, Connor, Upton 2007, p. 4), para identificar los elementos que subyacen al discurso, además de subrayar su coherencia – desde un punto de vista cognitivo – reflejada por la estructura del texto. Asimismo, el análisis cualitativo excluiría el cuarto punto de los problemas planteados por Cheng que se han enlistado antes.

En tercer, y último, lugar, se puede afirmar que el presente estudio tiene un contexto de producción muy claro. A partir de un artículo de 2020, periodo prepandemia, en España se analiza cómo algunas páginas web de dos museos (El Prado y Reina Sofía) se dirigen a su comunidad de referencia, influenciándose mutuamente. Por tanto, aunque no sea tan fácil establecer las fronteras de la mencionada comunidad de referencia, parece correcto sostener la precisión y la claridad de los criterios de elección del corpus – que es un asunto fundamental antes de empezar cualquier análisis lingüístico. En resumen: un corpus formado por las páginas web de los más importantes museos españoles según un artículo de “El País” de 2020, recogido con el fin de averiguar cómo estos sitios se relacionan con su comunidad de referencia.

Para alcanzar dicho objetivo, fue necesario también seleccionar las páginas para analizar. Un sitio internet es formado por varias secciones, que tienen su jerarquía interna y un preciso orden. Con el declarado fin de encontrar textos que explicitasen, por un lado, la auto-representación del museo y, por otro, la relación con la comunidad de referencia, se decidió reducir el estudio a tres secciones: *Historia*, *Misión* y *Educación*. En estas páginas se encuentran las estrategias que satisfacen los criterios explicitados anteriormente y que podrían representar lo que Biber, Connor y Upton definen como ‘move’: es decir, “a section of a text that performs a specific communicative function” (Biber, Connor, Upton 2007, p. 24). Por eso, se puede afirmar también que el presente estudio es primariamente inductivo: de hecho, el dato – es decir, los textos – ha guiado el análisis, aun empezando por el citado artículo de “El País”, que fue un primer recurso útil para reducir el corpus.

Una vez establecido todo eso, cabe examinar el segundo problema que se ha planteado: el método de análisis.

2.1. Metodología: el análisis del discurso

Los criterios del corpus analizado deben acompañarse por un método que pueda ayudar el desarrollo del estudio. En el párrafo anterior ya se ha especificado que la indagación fue primariamente inductiva y cualitativa. A estos dos conceptos parece necesario añadir algunos elementos.

En el presente trabajo se examina el discurso digital de los sitios de museos españoles. Eso significa, por lo tanto, adentrarse en el análisis del discurso, que sería la manera en la cual “people build and manage their social world using various semiotic systems” (Bou Franch, Garcés-Conejos Blivitch 2019, p. 4). El discurso sería, entonces, una práctica social, un recurso a través del cual los seres humanos interactúan construyendo y reflejando su propio mundo. Siguiendo este marco conceptual, se podría por ende afirmar que el discurso tiene tres funciones principales, según propone Fairclough: las “identity, relational and ideational functions” (Fairclough 1992, p. 64). Estas funciones están presentes en cualquier discurso (siempre en opinión de Fairclough), permiten construir identidades sociales (*identity*), relaciones sociales (*relational*) y un sistema de ideas (*ideational*). Todo lo expuesto hasta ahora cabe perfectamente en nuestro análisis: el discurso de las páginas web permite construir una identidad del museo, que se auto-representa; consiente elaborar relaciones sociales, en el momento en el que se conecta con la comunidad exterior; y, al final, trasmite un sistema de ideas, que representaría los pilares conceptuales del museo – es decir, su características fundacionales.

El análisis del discurso – cuyo objetivo es indagar la manera en la cual un texto se construye, por qué se produce y la razón de su estructura (Johnstone 2008, p. 9) – representa entonces la “transdisciplina” (Van Dijk 2016, p. 22) que mejor puede ayudar el desarrollo del presente estudio. En otras palabras, la indagación del texto, de sus herramientas, de sus estructuras, permitirá desvelar las relaciones que el mismo texto construye y refleja.

Eso puede afirmarse precisamente porque “discourse analysis sees identity practice as essentially discursive and relational, that is, as socially constructed. Identities emerge in interaction with others” (Bou Franch, Garcés-Conejos Blivitch 2019, p. 10).

Esta interacción entre discurso e identidad permanece en el discurso digital que, en palabras de Bou Franch y Garcés-Conejos Blivitch (2019, p. 11), puede también juntarse con el análisis crítico del discurso, precisamente por la capacidad de este último de desvelar la mutua influencia entre el texto y su mundo exterior. Empero, cabe detectar que uno de los pilares básicos del análisis crítico del discurso, es decir, el discurso de la élite y su ideología, no forma parte del presente trabajo – siendo la relación entre internauta y sitio no necesariamente *top-down*. Dicho de otro modo, el discurso museístico no supone una ideología que hay que desvelar, dado que no representa una élite social, sino una institución, cuyo objetivo es ofrecer una representación a su comunidad de referencia relacionada con su colección y su historia. Para averiguar eso, a parte del análisis de tipo lingüístico, es necesario un estudio del discurso que se centre en “the processes surrounding the production and consumption of texts” (Biber, Connor, Upton 2007, p. 6), identificando la

razón por la cual un texto se escribe y se utiliza (por el internauta) de una determinada manera. No necesariamente, dentro de este proceso, hay un desequilibrio social entre emisor y receptor del mensaje.

En conclusión, se puede defender la posición según la cual el análisis crítico del discurso, por su naturaleza, es útil al desarrollo de este estudio: sin embargo, no se tomará de manera integral, sino se adaptarán sus herramientas a cada caso analizado.

Después de haber introducido este marco teórico formado por la definición y las características de discurso digital, la definición del corpus y la metodología del análisis, hace falta solamente la comparación entre los dos museos: es decir, Prado y Reina Sofía.

3. La comparación: el Museo del Prado

Abierto al público en el noviembre de 1819 el Museo Nacional del Prado, en Madrid, representa uno de los más importantes patrimonios culturales de España y del mundo. En su historia han ingresado más de dos mil trescientas pinturas, además de una gran cantidad de esculturas, estampas y dibujos. En su interior, hoy en día se hallan obras de El Greco, Tintoretto, Mantegna, Goya, Velázquez y muchos otros artistas de fama mundial.

Su historia representaría, entonces, su misma identidad. Los puntos principales sobre los cuales el Prado hace hincapié en su auto-representación serían precisamente su importancia cultural dentro y fuera de la península ibérica, en calidad de institución secular que acompaña la historia de España.

Historia del museo

El edificio que hoy sirve de sede al Museo Nacional del Prado fue diseñado por el arquitecto Juan de Villanueva en 1785, como Gabinete de Ciencias Naturales, por orden de Carlos III. No obstante, el destino final de esta construcción no estaría claro hasta que su nieto Fernando VII, impulsado por su esposa la reina María Isabel de Braganza, tomó la decisión de destinar este edificio a la creación de un Real Museo de Pinturas y Esculturas. El Real Museo, que pasaría pronto a denominarse Museo Nacional de Pintura y Escultura y posteriormente Museo Nacional del Prado, abrió por primera vez al público en 1819 (del web del Prado¹²).

No es casual que la sección *Historia* se abra con estas palabras: hay referencias inmediatas a monarcas del siglo XVI, además de mencionar Juan de Villanueva, cuya figura fue fundamental para el desarrollo de la arquitectura neoclásica en España. No cabe utilizar un lenguaje afectado,

¹² <https://www.museodelprado.es/museo/historia-del-museo> (22.04.23).

nuestra (del Prado) historia ‘habla’ por sí misma, subrayando por tanto la importancia del museo. Una institución que posee “valiosísimas Colecciones Reales” que “comenzaron a tomar forma en el siglo XVI bajo los auspicios del emperador Carlos V y fueron sucesivamente enriquecidas por todos los monarcas que le sucedieron, tanto Austrias como Borbones” (del web del Prado¹³). En otras palabras, nuestro desarrollo es el desarrollo de España, su historia, a través de nuestras colecciones valiosísimas (nótese el uso del superlativo que acompaña las mayúsculas de respeto, para enfatizar el valor y la importancia del museo).

Siempre en la misma sección, se nota un preciso empleo de los enlaces (en **negrita** y en *cursiva* en el siguiente texto).

A ellos se deben los tesoros más emblemáticos que se pueden contemplar hoy en el Prado, tales como ***El jardín de las Delicias*** de **El Bosco**, ***El caballero de la mano en el pecho*** de **El Greco**, ***El tránsito de la Virgen*** de **Mantegna**, ***La Sagrada Familia conocida como La Perla*** de **Rafael**, ***Carlos V en Mühlberg*** de **Tiziano**, ***El Lavatorio*** de **Tintoretto**, el ***Autorretrato*** de **Durero**, ***Las Meninas*** de **Velázquez**, ***Las tres Gracias*** de **Rubens** o ***La familia de Carlos IV*** de **Goya** (del web del Prado, **negrita** y *cursiva* presentes en el original¹⁴).

Las partes en **negrita** y *cursiva*, si se hace clic, permiten acceder a la descripción de las obras o del autor, con informaciones biográficas y, también, un listado de las creaciones de un particular pintor poseídas por el museo. Es un claro ejemplo de aquella intertextualidad de que se ha tratado antes, y de la presencia del internauta que puede informarse sobre la importancia del museo en relación con las obras que se hallan en el edificio.

Las exposiciones están al centro de la atención también en la página principal del museo.

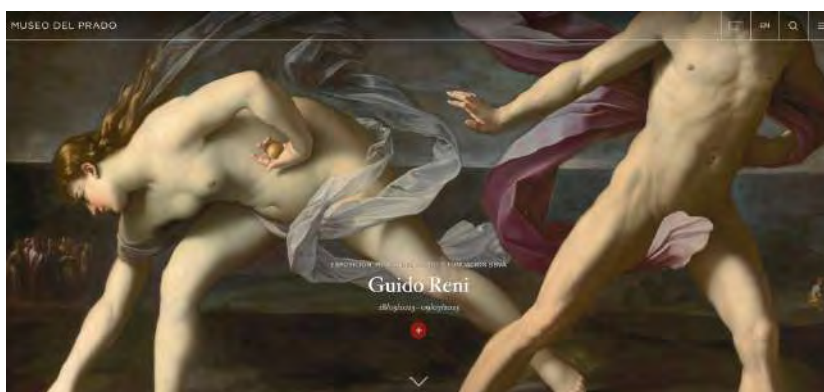


Imagen 2
La página principal del museo del Prado¹⁵.

¹³ <https://www.museodelprado.es/museo/historia-del-museo> (22.04.23).

¹⁴ <https://www.museodelprado.es/museo/historia-del-museo> (22.04.23).

¹⁵ <https://www.museodelprado.es/> (22.24.23).

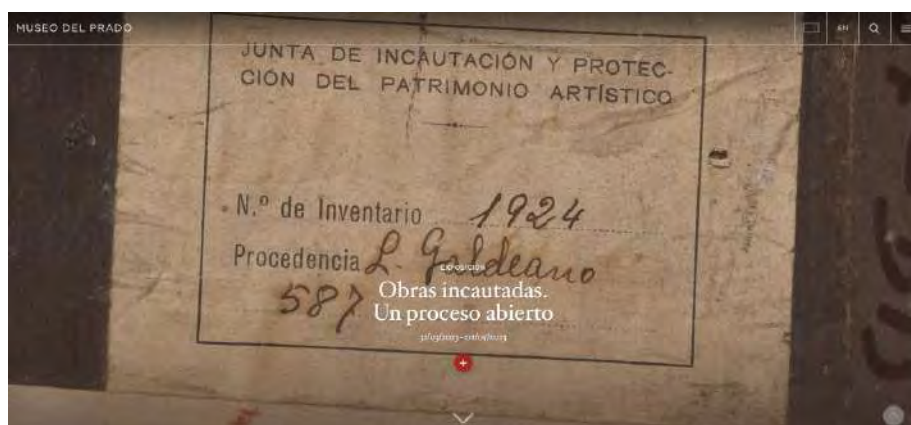


Imagen 3

La página principal del museo del Prado¹⁶.

La disposición de texto e imagen confirma lo dicho. El nombre de la institución se pone en alto a la izquierda, mientras que la imagen del trasfondo cautiva la atención del internauta, que puede hacer clic sobre el texto para profundizar la exposición que le parece más interesante. Nótese también la disposición de las imágenes: para pasar de una exposición a otra es necesario pulsar la flecha que permite ir abajo en la página. Eso significa que, por razones de importancia o de cronología, es el museo que ofrece un orden de navegación dentro de las muestras que ocurren sincrónicamente.

Además, como confirmación de la dimensión representativa e internacional del Prado, los idiomas a través de los cuales se puede navegar por el sitio son castellano e inglés. No hay espacio para las lenguas cooficiales de España: una precisa polarización que reafirma el papel nacional de museo. Al ser el Prado una institución histórica madrileña, no necesita incluir idiomas cooficiales para afirmar su personalidad: por el contrario, si lo hiciera quizá perdería su status. Aquí estriba la polarización, es decir, elegir una precisa posición dentro de dos polos normalmente opuestos¹⁷.

Como ya se ha expuesto, las secciones que más ayudan a entender la auto-representación del museo en relación con su comunidad de referencias son *Misión* y *Educación*. En el caso del Prado, no existe la sección llamada *Misión* que, en cambio, está en muchos otros sitios de museos. Se podría deducir que esta idea de misión se relaciona más con un concepto moderno

¹⁶ <https://www.museodelprado.es/> (22.24.23).

¹⁷ El concepto de polarización, en la publicidad, tiene que ver con el sector de producción de una empresa y la consiguiente elección de los estereotipos relacionados con el sector mismo: por ejemplo, lo natural o lo auténtico en el ámbito de la alimentación (Ferraz Martínez 2011, pp. 45-47). En el presente trabajo, este concepto se ha trasladado a la auto-representación del museo y a su imagen.

de empresa, como especifica el web del banco Santander: “La misión de una empresa se define como la razón principal por la cual esta existe, es decir, cuál es su propósito u objetivo y cuál es su función dentro de la sociedad” (del web de Santander¹⁸). Es un concepto que viene del inglés *mission*, entendido como “particular work that you feel it is your duty to do” (Hornby 2005, p. 978) y aplicado especialmente al mundo de los negocios.

Una noción de este tipo parecería chocar con la sección *Historia* analizada antes. Por ende, el sitio del Prado se presenta coherente con sus valores de institución histórica que ha contribuido al desarrollo de España, acompañando el país. Y cualquier internauta que busca dicha institución puede relacionarse con ella a través de la sección *Educación*.

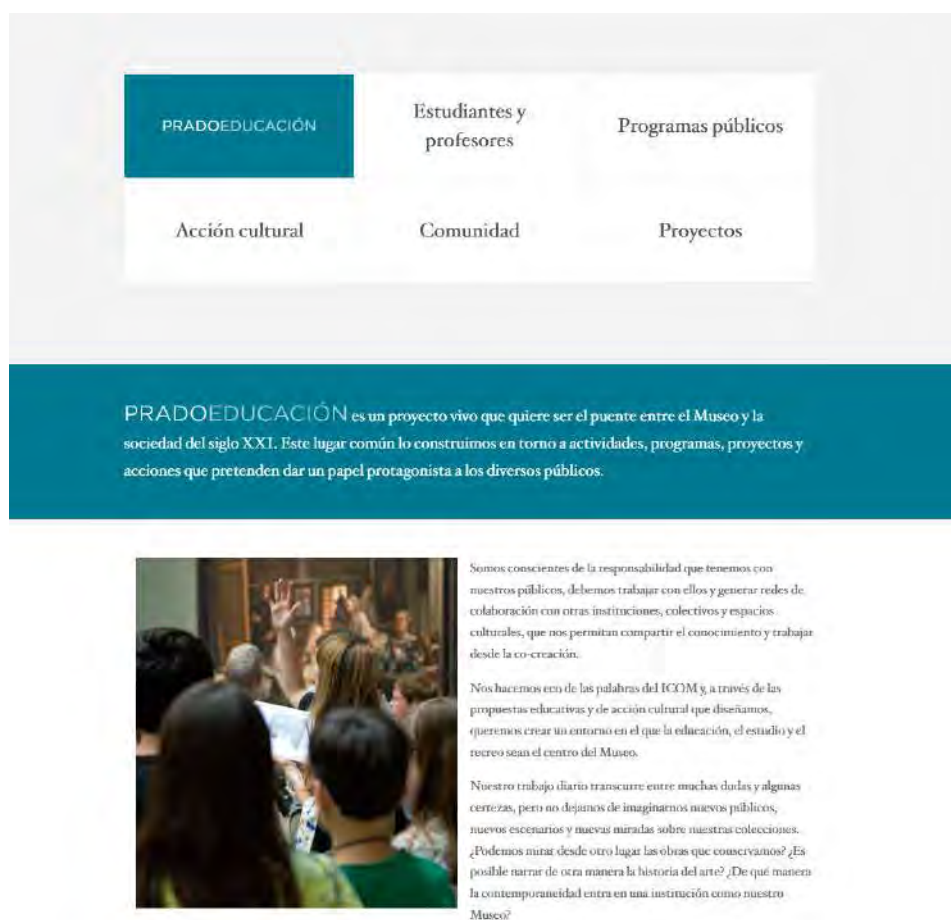


Imagen 4
Web del Museo del Prado, sección *Educación*¹⁹.

En alto pueden verse todas las sub-secciones a las cuales es posible acceder a través de los enlaces. Más abajo, un pequeño texto explica de manera clara el objetivo del proyecto *Prado educación*. Antes de analizarlo, cabe notar el uso

¹⁸ <https://www.becas-santander.com/es/blog/mision-vision-y-valores.html> (22.04.23).

¹⁹ <https://www.museodelprado.es/aprende/pradoeducacion> (25.04.23).

de las mayúsculas y de las negritas que subrayan el nombre *Prado educación*: este recurso gráfico se utiliza mucho, tanto en el web como en la publicidad, para que el receptor del mensaje ponga la atención en uno o más elementos. En este caso, el título del proyecto, que es también el título de la sección, es la primera componente para evidenciar. Luego, el texto aclara objetivos y posibles destinatarios de *Prado educación*, que es:

[...] un proyecto vivo que quiere ser el puente entre el Museo y la sociedad del siglo XXI. Este lugar común lo construimos en torno a actividades, programas, proyectos y acciones que pretenden dar un papel protagonista a los diversos públicos (del web del Prado²⁰).

El proyecto sirve para conectar, exactamente como un puente (utilizando, entonces, una clara metáfora espacial), museo y sociedad en el siglo XXI. Un “lugar común” que “construimos” – el deíctico personal nosotros, incluido en el verbo, evidencia la cooperación entre los sujetos participantes – en torno a varias iniciativas, útiles para que “los públicos”, que juntos formarían la comunidad de referencia, puedan ser protagonistas. Una conexión, por ende, que se realiza a través de la cooperación con la sociedad del siglo XXI: una posible inferencia sería que, a pesar de nuestra historia, como museo no estamos encima de la sociedad, sino colaboramos con ella, compartiendo nuestro tesoro cultural. Estos conceptos – importancia histórica, cooperación y compartición de nuestro patrimonio – se exaltan todavía más en las primeras líneas de la sección *Educación*:

Somos conscientes de la responsabilidad que tenemos con nuestros públicos, debemos trabajar con ellos y generar redes de colaboración con otras instituciones, colectivos y espacios culturales, que nos permitan compartir el conocimiento y trabajar desde la co-creación (del sitio del Prado²¹).

La “responsabilidad” viene de la grande riqueza cultural e histórica del museo que, además, es la razón por la cual “generar redes de colaboración” que permitan compartir el conocimiento. Nótese, al final del párrafo, la idea de la “co-creación”: trabajar en siglo XXI, según el museo del Prado, significa colaborar con una comunidad plural, formada por “instituciones, colectivos y espacios culturales”. Esta comunidad es protagonista, porque participa activamente en una creación compartida²². Por lo tanto, como se

²⁰ <https://www.museodelprado.es/aprende/pradoeducacion> (25.04.23).

²¹ <https://www.museodelprado.es/aprende/pradoeducacion> (25.04.23).

²² Este concepto de cooperación es analizado desde una perspectiva semiótico-textual por Umberto Eco que sostiene que “un testo vuole che qualcuno lo aiuti a funzionare” (Eco: 2020, p. 72). En otros términos, para que un texto adquiera significado necesita de una cooperación entre el autor – que lleva concientemente “interstizi da riempire” – y un lector que llena los vacíos del texto mismo.

afirmaba al principio de estas páginas, no es fácil establecer las fronteras de dicha comunidad de referencia, al ser esta multifacética ante todo según el emisor del mensaje. Lo único que se puede intuir es la idea que el museo tiene del receptor del mensaje: una comunidad cultural – escuelas, por ejemplo – que se interesa al tesoro del Prado y permite divulgarlo para co-crear conocimiento.

Todo lo que se ha afirmado hasta ahora se subraya más abajo en la misma página, donde se detallan más la historia del proyecto Educación y a quién se dirige.



El Área de Educación del Museo Nacional del Prado nace en 1983 como una necesidad para conectar las colecciones con la sociedad. A lo largo de estos más de treinta años, ha ido creciendo y modificando sus propuestas, actividades y metodologías para adaptarse a los cambios sociales y a las nuevas demandas de los públicos.

Nos sabemos responsables de cuidar, investigar y difundir un valioso patrimonio que nos pertenece a todos, y por lo tanto nuestras acciones trascienden lo meramente artístico para poner el acento en valores universales como el respeto, la igualdad, la interculturalidad o la sostenibilidad.

Así, de manera progresiva, se han establecido cuatro líneas de acción: dentro de la educación formal Programas Educativos dirigidos a **estudiantes y profesores**, desde la educación no formal los denominados **Programas Públicos** para jóvenes, familias, público infantil y adulto; **Comunidad** que recoge nuestra acción social y el programa de **Acción Cultural** con propuestas que parten de la relación de la pintura y la escultura con otros lenguajes artísticos como la música, la danza, el teatro o la performance.

Para llevar a cabo nuestra función de difusión del conocimiento nos servimos de diferentes herramientas y recursos educativos como publicaciones didácticas, contenidos online, exposiciones didácticas, materiales accesibles, aplicaciones, web, etc.



Imagen 5

Sitio del Museo del Prado, sección *Educación*²³.

Ya a partir del título, “Nuestra historia en el museo” se especifica la idea de comunidad a través del deíctico personal. Dicha historia sería al mismo tiempo la de España y la que se puede co-crear. Después de una breve introducción sobre el nacimiento del proyecto *Educación*, se pone énfasis

²³ <https://www.museodelprado.es/aprende/pradoeducacion> (25.04.23).

inmediato en dos elementos ya tratados antes, es decir, el patrimonio del Prado y los elementos que forman su comunidad imaginada²⁴.

Nos sabemos responsables de cuidar, investigar y difundir un valioso patrimonio que nos pertenece a todos, y por lo tanto nuestras acciones trascienden lo meramente artístico para poner el acento en valores universales como el respeto, la igualdad, la interculturalidad o la sostenibilidad.

Así, de manera progresiva, se han establecido cuatro líneas de acción: dentro de la educación formal Programas Educativos dirigidos a **estudiantes y profesores**, desde la educación no formal los denominados **Programas Públicos** para jóvenes, familias, público infantil y adulto; **Comunidad** que recoge nuestra acción social y el programa de **Acción Cultural** con propuestas que parten de la relación de la pintura y la escultura con otros lenguajes artísticos como la música, la danza, el teatro o la performance.

Para llevar a cabo nuestra función de difusión del conocimiento nos servimos de diferentes herramientas y recursos educativos como publicaciones didácticas, contenidos online, exposiciones didácticas, materiales accesibles, aplicaciones, web, etc. (del web del Prado, negrita presente en el original²⁵).

El texto empieza repitiendo el papel de responsabilidad del museo, fundamental en relación con su historia y su patrimonio – que es *nuestro*, es decir, de todos los españoles. En esta primera parte, resulta interesante notar cómo dicho patrimonio, y su gestión, conduce a valores universales y modernos como respeto, igualdad, interculturalidad y sostenibilidad. Sobre todo los últimos dos reafirman, en cierto modo, la conexión con el siglo XXI: se retoma así la idea inicial del proyecto educación, que sería la de construir un puente entre museo y sociedad en la actualidad. Por eso, se ponen en negrita todas las instituciones que forman parte de la comunidad de referencia (ahora un poco más que imaginada): estudiantes y profesores; quién quiera suscribirse a programas públicos para una educación no formal (“Programas Públicos”); personas que forman parte de asociaciones para ayudar quién tenga necesidades específicas (“Comunidad”) y, finalmente, otras asociaciones de danza, música o teatro (“Acción cultural”, que subraya la apertura del museo a relaciones intersemióticas). Abriendo los enlaces en negrita es posible obtener mayor información y suscribirse a un específico programa.

²⁴ El concepto de comunidad imaginada fue introducido, en el debate sobre la formación de la nación y del nacionalismo, por Benedict Anderson en 1983 (Anderson: 2006). Anderson consideraba que cualquier forma de nacionalismo se basa en principios compartidos, y que en el proceso de formación de estos principios la prensa desempeña un papel decisivo. El concepto, en la acepción del presente trabajo, no tiene los rasgos de unión de tipo nacionalista, sino se considera como un grupo de personas que se juntan bajo algunas ideas comunes – en este caso, la existencia del Museo del Prado y de su tesoro cultural.

²⁵ <https://www.museodelprado.es/aprende/pradoeducacion> (25.04.23).

En conclusión, es posible notar cómo las imágenes refuerzan la comprensión del texto, siendo en plena consonancia con los conceptos expresados: con una inferencia de tipo semiótico-pragmático, podríamos sostener que la señora mayor que lee la información sobre una determinada obra de arte sería parte de quién decide incorporarse a la comunidad mediante programas públicos. Además, los chicos que, en frente de *Las meninas*, estudian, representan aquel grupo de estudiantes que pueden enriquecer la comunidad del Prado, difundiendo sus valores modernos a partir de una obra cumbre de la pintura española y mundial. Aquel puente entre museo, pasado y presente que el Prado quiere representar de este modo toma forma.

Como conclusión de este breve análisis, se pueden evidenciar algunos aspectos que resumimos a continuación:

1. La importancia de la historia: el Prado es un museo que acompaña la vida de los españoles, dado que nació hace dos siglos y tiene obras de valor inestimables.
2. El puente entre pasado y presente: que permite conectar valores y personas, en una comunidad multifacética que co-crea la identidad del museo.
3. Cierta apertura del museo, que solo en apariencia choca con su identidad clásica. La convivencia entre estos elementos (arte clásico, historia y modernidad) parecería representar el trabajo más elaborado de la auto-representación del Museo del Prado.

4. La comparación: el museo Reina Sofía

Aunque el nacimiento del museo Reina Sofía es un evento bastante reciente, al remontar a 1990, en realidad la historia de su edificio es mucho más larga, dado que empezó en el siglo XVI con Felipe II. En origen, dicho edificio era un hospital.

La importancia del Reina Sofía se debe a su colección de arte moderno y contemporáneo, que incluye trabajos de Dalí, Miró y José Victoriano González-Pérez (conocido como Juan Gris), además del célebre *Guernica*, obra cumbre de Picasso con un significado histórico de primaria importancia.

Estos dos aspectos – la historia del edificio y la identidad del museo – conviven en la sección *Historia* del sitio, aunque de manera distinta con respecto al Prado. En el caso de este último, historia y edificio concurren a la auto-representación del museo; mientras que para el Reina Sofía parecen ser dos elementos distintos. Esto no significa que el uno pierda importancia con respecto al otro: pero, es imposible aquella confluencia que es parte esencial del Prado.

La apertura del Museo Nacional Centro de Arte Reina Sofía en 1990 supuso la creación de un museo de arte moderno y contemporáneo en España de nivel internacional, aunque no han sido pocos los avatares sufridos por el edificio hasta alcanzar tal fin.

La primera fundación del Hospital San Carlos, actual sede del Museo, se debe al rey Felipe II, quien, en el siglo XVI, centralizó en este lugar todos los hospitales que estaban dispersos en la Corte. En el siglo XVIII, Carlos III decidió una nueva fundación, al resultar insuficientes las instalaciones para la ciudad. El actual edificio es obra de los arquitectos José de Herosilla y Francisco Sabatini, debiéndose a este último gran parte de la obra.

En 1788, tras la paralización de las obras con la muerte de Carlos III, el edificio fue ocupado para asumir la función para la que había sido construido, la de Hospital, aunque lo edificado no fuera sino sólo un tercio del proyecto de Sabatini.

Desde entonces sufrió varias modificaciones y añadidos hasta que, en el año 1965, se clausura el hospital, pasando sus funciones a ser desempeñadas por la Ciudad Sanitaria Provincial. Logra sobrevivir a diversos rumores de demolición y, por medio de un real decreto de 1977, es declarado Monumento Histórico-Artístico, garantizando así su supervivencia (del web del museo Reina Sofía²⁶).

Desde el principio se afirma con claridad la identidad del museo: de nivel internacional y de arte moderno y contemporáneo. Solo luego se trata de los “avatares” del edificio, hasta llegar a la declaración de 1977 que oficializa el estatus de “Monumento Histórico-Artístico” por parte del gobierno. La historia del museo, en este caso, no parece acompañar la historia del país como para el Prado: la inferencia que puede producir el receptor sería la de un lugar histórico que se ha transformado en artístico con el fin de garantizar su supervivencia. Abajo en la página web hay también una galería fotográfica que incluye el folleto de inauguración del museo, explotando el potencial multimodal del web.

A diferencia del Prado, el sitio se presenta en cinco idiomas: castellano, inglés, catalán, gallego y euskera. Una atención a las diferencias internas a España que refuerza la idea de un museo inclusivo, también con respecto a la historia de las Comunidades Autónomas.

Como ya se había adelantado, otra diferencia con el Prado es la presencia de la sección llamada *Misión*. Esta parte se compone de un texto muy largo que empieza explicando los problemas que puede tener un museo en calidad de institución pública.

Aunque el Museo como institución pública aún mantiene su importancia en la trama de industrias creativas, en la actualidad ha perdido una parte de su poder de mediación o, al menos su posición privilegiada, en la definición de lo que hoy entendemos por cultura.

²⁶ <https://www.museoreinasofia.es/museo/historia> (25.04.23).

Quienes conforman el panorama cultural son, por un lado, los grandes actores de la industria de la cultura y de la comunicación y, por otro, el magma difuso de los productores que actúan desde la subordinación de su singularidad creativa, ya sea vendiendo su capacidad de creación, o siendo expropiados de ella. Además nos hallamos inmersos en una profunda crisis del sistema de la que el museo no es ajeno. Si el paradigma económico basado en la especulación y el dinero fácil no se sustenta, es también evidente que la primacía del edificio y del espectáculo sobre el programa artístico del museo ha dejado de tener validez. La exigencia de inventar otros modelos es imperiosa.

La generación de un nuevo modelo cultural debe ir acompañada de cambios institucionales, porque las instituciones son las principales estructuras de invención de lo social, de un hacer afirmativo y no limitativo. Esto es más importante en nuestra época porque en la sociedad moderna occidental las artes de gobierno no consisten en aplicar medidas represivas, sino en hacer que éstas se interioricen. Con la llegada en las últimas décadas de lo que Luc Boltanski y Ève Chiapello denominaron crítica artística como forma característica de las relaciones laborales, el individuo pasa, de forma natural y no forzada, a jugar un papel activo en su propia sujeción al dominio gubernamental. La crítica artística reclamaba una vida auténtica, no alienante, asentada en la creatividad y en la no dependencia de un patrón y unos horarios fijos prototípicos del fordismo. Pero a la vez, promovía la subordinación del sujeto a una estructura laboral en la que es el productor cultural mismo quien favorece su propia precarización. Este último intenta alcanzar una mayor libertad y flexibilidad a costa de la expropiación de su trabajo a manos de quienes poseen el capital o las vías legales de desposesión. O lo que es lo mismo, a costa de la introducción de su propia actividad creativa en la lógica del mercado o en formas de dominio cultural al servicio de proyectos de apropiación del espacio público (del web del museo Reina Sofía²⁷).

En esta introducción se saca una especie de fotografía de la situación artística del país y, por consiguiente, de las instituciones que se ocupan de arte y cultura. La “industria de la comunicación” y el “magma difuso de los productores” son los actores que priman en este sector: todo ello supone “la exigencia de inventar otros modelos” para salir de una situación difícil, que es dependiente de las leyes del mercado. En efecto, un poco más adelante en el texto se reafirma lo dicho especificando mejor lo que está ocurriendo.

[...] es esencial romper la dinámica de franquicias, que tanto parece atraer a los responsables de los museos, y pensar más bien en una especie de archivo de lo común, de una confederación de instituciones que compartan las obras que albergan sus centros y, sobre todo, participar las experiencias y relatos que se generan a su alrededor. Sólo así podremos decir que poner el yo en plural depende de mi implicación en el mundo con los demás, y no en mi acceso al otro. Es en ese lugar, emplazado entre el yo y el otro, donde se realiza la esfera de lo común. Una esfera que es distinta de la pública. Lo público, en el fondo,

²⁷ <https://www.museoreinasofia.es/museo/mision> (25.04.23).

no nos pertenece. Lo público que nos proporciona el Estado reside meramente en la gestión económica delegada por un todo colectivo en la clase política. Lo común no es una expansión amplificada de lo individual. Lo común es algo que nunca se lleva a término. Lo común sólo se desarrolla a través del otro y por el otro, en la sede común, en el ser compartido, por utilizar los términos de Blanchot (del sitio del museo Reina Sofía²⁸).

Por ende, el pasaje del “público” a lo “común” resulta fundamental, y parece ser la piedra angular del discurso del Reina Sofía: un pasaje que supone un cambio de paradigma. De hecho, lo público tendría sobre todo que ver con la gestión económica del Estado, mientras que lo común es un proceso que se desarrolla entre individuos, “del otro y por el otro”, en un sinfín de relaciones. Se cita Maurice Blanchot, filósofo y escritor surrealista, para reforzar el concepto explotando por un lado el poder del metadiscurso²⁹, y por otro implícitamente reafirmando la naturaleza de museo de arte moderno – siendo el surrealismo una corriente del siglo XX.

En definitiva, la misión del museo sería la de cambiar este paradigma, transformándose en un museo de lo común a través de algunos conceptos básicos que representan la identidad de la institución.

[...] Desde el Museo Reina Sofía estamos desarrollando varias líneas que buscan precisamente la transformación del museo en un museo de lo común:

1. La **Colección**. [...] Propugnamos una identidad relacional que no es única ni atávica, sino de raíz múltiple. Esta situación determina la apertura al otro y plantea la presencia de otras culturas y modos de hacer en nuestras propias prácticas, sin miedo a un hipotético peligro de disolución. [...] La relación no va de lo particular a lo general, o viceversa, sino de lo local a la totalidad-mundo, que no es una realidad universal y homogénea, sino plural.
2. Trabajamos en la creación de un archivo de lo común. Una especie de **archivo de archivos** [...] Derrida nos explica que el archivo es a la vez un topos, un lugar y un nomos, la ley que lo organiza. En el archivo de lo común esta ley es compartida, no instituida, sino instituyente. No responde a una genealogía del poder, ni ordena jerárquicamente los saberes de la sociedad.
3. Finalmente, desde el Reina Sofía se está organizando una red heterogénea de **trabajo con colectivos**, movimientos sociales, universidades, etc. que cuestionan el museo y que generan ámbitos de negociación no meramente representativos (del web del museo Reina Sofía. Negrita presente en el original³⁰).

²⁸ <https://www.museoreinasofia.es/museo/mision> (28.04.23).

²⁹ El concepto de metadiscurso fue estudiado por Norman Fairclough, que lo considera como un recurso útil para corroborar el propio punto de vista. Al utilizar esta herramienta, es común citar fuentes consideradas acreditadas como instituciones, informes o autores de primaria importancia, para fortalecer una determinada argumentación (Fairclough 1992).

³⁰ <https://www.museoreinasofia.es/museo/mision> (28.04.23).

Estos tres puntos (colección, archivo de los archivos y trabajo con colectivos) subrayados en negrita para facilitar su individuación, son la misión del museo. Su personalidad, sus objetivos.

La estrategia discursiva es clara y totalmente distinta con respecto a lo visto con el Museo del Prado. Toda esta sección *Misión* se presenta con algunas características que otorgan mucha fuerza al texto y fijan la identidad de la institución.

1. **Léxico afectado**: hay una precisa elección terminológica que confiere al texto una fuerte dimensión intelectual. Es un escrito que parece dirigido a personas que trabajan en el sector museístico, que ya conocen la situación aunque solo de manera superficial. Piénsese a sintagmas como “raíz múltiple”, “apertura al otro”, “otras culturas”, “archivo de lo común”, “red heterogénea”, que subrayan también la naturaleza compartida de la relación entre el museo y su comunidad de referencia. Cabe notar cómo dicha naturaleza no se forma a través de un puente, como en el caso del Prado, sino a través del archivo de lo común, para cambiar un paradigma que, según la sección *Misión* del Reina Sofía, ya no funciona.
2. **Metadiscurso**: siguiendo la idea del texto intelectual, para fortalecer la propia argumentación el autor menciona también a Jacques Derrida, importante filósofo cuyo nombre se conecta principalmente con el concepto de deconstrucción. La obra que se cita es *Mal de archivo*, publicado en 1995 donde se define el archivo como “un topos, un lugar y un nomos”. El metadiscurso, en este caso, sirve entonces para sostener uno de los pilares de la identidad del museo, es decir, el archivo de lo común – compartido y no jerárquicamente ordenado.
3. **Idea de relación abierta entre museo y comunidad de referencia**: coherentemente con lo dicho hasta ahora, el Reina Sofía se conecta con una red heterogénea de sujetos que pueden incluso cuestionar el museo. Un aspecto, este último, totalmente ausente en el Prado: hay sí relaciones, pero no se llega al punto de cuestionar la institución de referencia.
4. **Coherencia discursiva**: leyendo el texto se percibe la idea de un museo abierto, moderno, con una alta dimensión intelectual. Todo llevado a cabo por una precisa elección lexical y una clara estrategia discursiva.

La última sección que cabe analizar es *Educación*. Su estudio puede confirmar una fuerte coherencia con la *Misión* del Reina Sofía, subrayando al mismo tiempo algunas diferencias con el Prado.

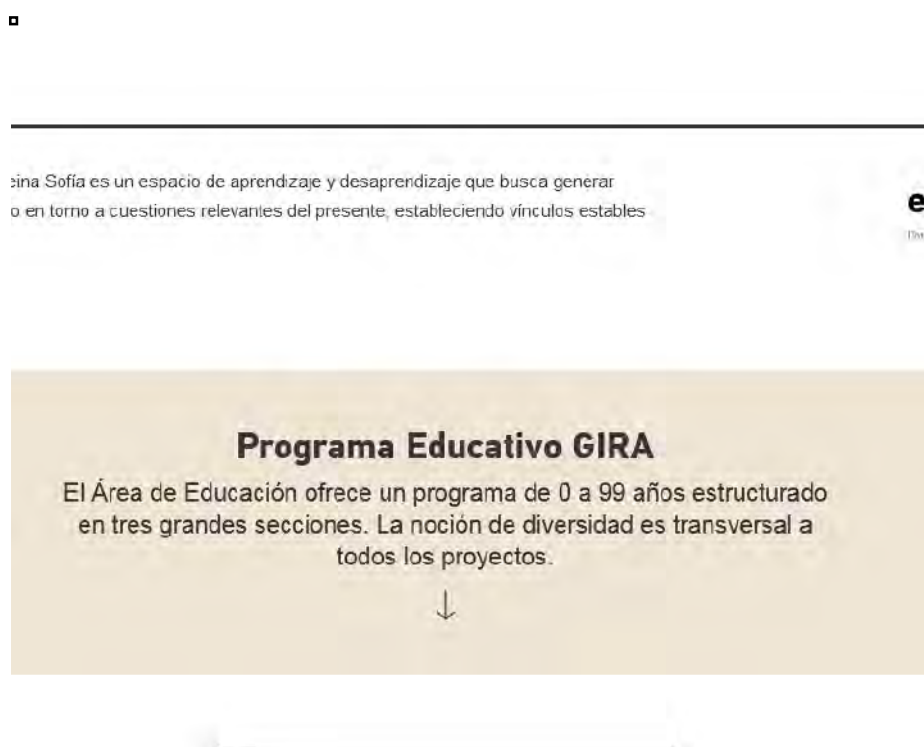


Imagen 6
Sitio del Museo de Reina Sofía, sección *Educación*³¹.

Abriendo la página inicial de esta sección aparecen dos textos distintos. El primero define de manera general la idea que subyace al concepto de *Educación*.

El Área de Educación del Museo Reina Sofía es un espacio de aprendizaje y desaprendizaje que busca generar experiencias de conocimiento crítico en torno a cuestiones relevantes del presente, estableciendo vínculos estables entre el Museo y sus públicos (del web del Prado³²).

La metáfora espacial – solo aparente, dado que el museo es efectivamente un espacio para mostrar obras – acompaña la definición de educación, llevada a cabo por una antonimia que se parece a un oxímoron: “aprendizaje y desaprendizaje”. Vuelve, por ende, el concepto de cuestionar la institución, un lugar donde se aprende también a desaprender, con el único fin del conocimiento crítico sobre asuntos del presente. Se podría por lo tanto afirmar que la comunidad imaginada por el Museo Reina Sofía es también epistémica: es decir, parafraseando a Van Dijk (2016, p. 16), un grupo de personas que comparte y reproduce el conocimiento. Y, además, lo cuestiona (siguiendo el enfoque del Reina Sofía): por lo tanto, es posible otra vez observar la fuerte coherencia con la sección *Misión*. Los “vínculos estables”

³¹ <https://www.museoreinasofia.es/educacion> (29.04.23).

³² <https://www.museoreinasofia.es/educacion> (29.04.23).

entre museo y sus públicos – nótese el plural que reafirma la pluralidad que dialoga con la institución – permiten la creación de esta comunidad epistémica que se relaciona con el museo a través del llamado proyecto educativo GIRA, que sería “un programa de 0 a 99 años estructurado en tres grandes secciones. La noción de diversidad es transversal a todos los proyectos” (del web del Reina Sofía³³).

En conclusión en este espacio compartido, donde todos pueden ayudar al museo y cuestionarlo si necesario, se puede indagar la realidad con una mirada crítica, gracias a los proyectos del Museo Reina Sofía que permiten aprender y desaprender en el llamado museo de lo común.

5. Conclusiones

Este breve análisis ha permitido comparar los sitios web de dos grandes museos madrileños, averiguando como su auto-representación difiere a partir de su personalidad. El discurso, sea este digital, escrito u oral, sigue siendo una herramienta fundacional de la identidad: es decir, un recurso de los seres humanos que consiente afirmar quiénes somos, nuestras relaciones, nuestros pilares básicos.

El web no representa una excepción en este sentido, aunque su formato muestra algunos rasgos que complican la indagación: multimodalidad, la presencia del internauta y la posibilidad de cambiar constantemente. Por un lado, todo eso parece casi reforzar la idea derridiana de deconstrucción del texto³⁴, de su imposibilidad de conservarse de manera definitiva. Por otro, y por consiguiente, se plantean algunos problemas que pueden considerarse típicos del análisis del discurso: parcialidad, recogida del material para definir el corpus, conservación de las páginas web, estudio cualitativo o cuantitativo. Sin embargo, si se define desde el principio el método de manera clara, la indagación del discurso museístico puede generar resultados muy interesantes, desvelando cómo una institución quiere representarse, con cuáles estrategias, con quién se relaciona – en una conexión bidireccional en la que los actores se influyen mutuamente.

En el caso específico, tanto el Prado como el Reina Sofía se apelan a valores modernos como pluralidad, inclusión, interculturalidad, sostenibilidad. La idea básica es la de instituciones que, a través del arte, evolucionan con el tiempo por su misma naturaleza: con todo, lo hacen de modo distinto, relacionándose con comunidades diferentes. Y el análisis del

³³ <https://www.museoreinasofia.es/educacion> (29.04.23).

³⁴ Una síntesis muy eficaz de las teorías de Derrida puede encontrarse en el volumen de Mario Vergani publicado en 2000.

discurso permite averiguar estas afirmaciones, como los textos estudiados reflejan y construyen la identidad de los museos.

No cabe duda que el presente estudio podría continuar y también extenderse, englobando otras instituciones – según el artículo de “El País” mencionado al principio de estas páginas. El objetivo permanecería el mismo, es decir, comparar identidades y comunidades a través del discurso, siempre a partir del dato y fijando claramente los pilares del análisis.

Bionote: Doctor en Estudios humanísticos (2019) por la Universidad de Módena y Reggio Emilia, es Profesor Asociado en la misma Universidad para el curso OFA (Español para principiantes), Departamento de Estudios Lingüísticos y Culturales. Trabaja también como profesor de lengua y cultura inglesa en el Instituto Lazzaro Spallanzani, Castelfranco Emilia, Modena. Impartió cursos de Español I y II en la facultad de Economía de Modena y Reggio Emilia, y de Cultura y Literatura española en la Universidad de Mantua. Sus intereses de investigación se centran en el análisis del discurso, con particular atención en el lenguaje del web y en las estrategias de auto representación de la identidad en procesos de *nation building* en España. En estos procesos, el estudio de la cultura taurina desarrolla un papel de primaria importancia.

Author’s address: simone.tepedino@unimore.it

Bibliografía

- Anderson B. 2006, *Imagines Communities. Reflections on the origin and spread of nationalism*, Verso, London.
- Biber D., Connor U., Upton T. 2007, *Discourse on the Move*, John Benjamin Publishing Company, Amsterdam/Philadelphia.
- Bou-Franch P., Garcés-Consejos Blitvich P. (ed.) 2019, *Analyzing digital discourse. New insights and future directions*, Palgrave Macmillan, Switzerland.
- Carliner S., Shi X., Wenjing W. 2020, *Internet-mediated genre studies: an integrative literature review (2005-2019)*, en “IEEE Transactions On Professional Communication” 63 [4], pp. 279-295.
- Calvi M.V. 2016, *Guía de viaje y turismo 2.0: los borrosos confines de un género*, en “Ibérica” 31, pp. 16-38.
- Cheng W. 2012, *Exploring Corpus Linguistics. Language in Action*, Routledge, Abingdon.
- Cortellazzo M. 2013, *Metodi qualitativi e quantitativi di analisi di testi*, en “Contemporanea” 16, pp. 299-300.
- Covadonga L.A. 2014, *Análisis del discurso*, Editorial Síntesis, Madrid.
- Eco U. 2020, *Lector in Fabula. La cooperazione interpretativa nei testi narrativi*, La nave di Teseo, Milano.
- Fairclough N. 1992, *Discourse and Social Change*, Polity Press, Cambridge.
- Ferraz Martínez A. 2011, *El lenguaje de la publicidad*, Arco Libros, Madrid.
- Gianmatteo M., Gubitosi P., Parini A. 2017, *El español en la red*, Iberoamericana Vervuert, Madrid, Frankfurt.
- Halliday M.A.K. 1989, *Spoken and written language*, Oxford University Press.
- Herring S.C. (ed.) 1996, *Computer Mediated Communication. Linguistic, Social and Cross-Cultural Perspectives*, John Benjamins Publishing, Amsterdam.
- Hornby A.S. 2005, *Oxford Advanced Learner’s Dictionary of Current English*, Oxford University Press, Oxford.
- Jakobson R. 1985, *Ensayos de lingüística general*, Planeta-Agostini, Barcelona.
- Johnstone B. 2008, *Discourse Analysis*, Blackwell Publishing, Malden.
- Kress G. 2003, *Literacy in the new Media Age*, Routledge, London.
- Lepkowska-White L., Imboden K. 2013, *Effective Design for Usability and Interaction: The Case of Art Museum Websites*, en “Journal of Internet Commerce” 12, pp. 284-305.
- López Pena Z. 2021, *Una propuesta para el análisis multimodal de dos sitios web oficiales turísticos en torno a la ciudad de Santiago de Compostela*, en “Discurso & Sociedad”, pp. 834-859.
- Mancera Rueda A., Pano Alamán A. 2013, *El español coloquial en las redes sociales*, Arco Libros, Madrid.
- Pallud J., Straub Detmar W. 2014, *Effective website design for experience-influenced environments: The case of high culture museums*, en “Information & Management” 51, pp. 359-373.
- Rainer W. (eds.) 1989, *Estética de la recepción*, La Balsa de la Medusa, Madrid.
- Van Dijk Teun A. 2016, *Discurso y conocimiento. Una introducción multidisciplinaria*, Editorial Ariel, Barcelona.
- Rojo G. 2021, *Introducción a la lingüística de corpus en español*, Routledge, London.
- Vergani M. 2000, *Jacques Derrida*, Bruno Mondadori, Torino.
- Volli U. 2003, *Manuale di Semiotica*, Editori Laterza, Roma.

Yu Meng 2021, *Interactive Design Of a Museum Website*, en Becker, Kupryianov, Metveev, Taratukhin (eds.), *Information System and Design. Second International Conference on Intelligent Design (ICID)*, Springer, Berlin, pp. 251-254.

PRESENCIA DEL LÉXICO COLOQUIAL EN LAS ENTREVISTAS DE JORDI ÉVOLE

Estudio de corpus

ANDRÉS ORTEGA GARRIDO
UNIVERSITÀ DI BERGAMO

Abstract – Jordi Évole, a Spanish journalist mainly known for his interview program *Salvados* –a broadcast by the La Sexta network– has revitalized the interview genre by promoting ease and naturalness. These characteristics appear both in the choice of the location for the interview and in the casual and relaxed approach to the interaction with the interviewee. From a linguistic point of view, the lexicon used by the journalist is strongly colloquial, sometimes far from the traditional formality of this type of communicative exchanges. The colloquial vocabulary used in the program is one of its keys to success, since it emphasizes the identification between the public and the journalist through a constant use of words and expressions of common speech, including interjections and a vulgar lexicon. Starting from a corpus formed by the transcription of a series of interviews from the *Salvados* program, we will quantitatively and qualitatively analyze the lexicon used by Évole with the methodological tools offered by Corpus Linguistics.

Keywords: colloquial lexicon; journalistic interview; *Salvados*; Évole; Corpus Linguistics.

1. La entrevista periodística desde el análisis de la conversación

La entrevista periodística, en su vertiente televisiva, se engloba en un tipo de discurso conversacional donde es posible detectar una asimetría en cuanto a la relación que se establece entre los interlocutores, de modo que tal intercambio comunicativo se construye de forma diferente a la conversación espontánea prototípica. Más concretamente, deberíamos adscribir la entrevista a las interacciones de tipo institucional (Orletti 2000), que normalmente tienen lugar en un marco físico bien definido, como el interrogatorio policial en comisaría, la vista oral de un juicio en un tribunal, la conversación entre médico y paciente en la consulta de este último o la interacción entre profesores y alumnos en las aulas de los centros escolares. La peculiaridad de las entrevistas periodísticas es que no se ubican en un escenario preestablecido. Es más, en el caso de las de Jordi Évole en el programa de televisión *Salvados*, emitido en la cadena española La Sexta (actualmente también disponible a través de la página web de Atresmedia) y

que presentó de 2008 a 2019, estas se desarrollan en espacios inusuales, como, por ejemplo, lugares públicos (parques, plazas, la calle, un bar...) o el interior de un coche, a diferencia de la tradicional entrevista en un estudio de televisión. A pesar de tales peculiaridades, este género del periodismo oral, incluido el modelo propuesto por el citado periodista, participa de la mayor parte de las características de la interacción institucional, razón por la cual consideramos que se puede estudiar como tal (véase Orletti 2000, pp. 73-74).

En efecto, *Salvados* se caracteriza por poner en práctica un tipo de entrevista que en muchos aspectos se distancia de los esquemas habituales. De hecho, es verosímil que uno de los objetivos del programa en el momento de su aparición fuera precisamente alejar este género periodístico de su tradicional rigidez. Con todo, estas entrevistas mantienen una serie de reglas y principios en función de factores como el espacio en que se producen y, sobre todo, del papel que desempeña cada uno de los intervinientes.

A este respecto, conviene subrayar que desde el punto de vista de la comunicación existe una asimetría evidente en cuanto a los roles prefijados de entrevistador y entrevistado, roles establecidos previamente y de manera tácita. Por otro lado, no hay que olvidar que en este tipo de intercambios comunicativos asimétricos existe el llamado ‘director de la interacción’ (Orletti 2000, pp. 18-26), encargado de guiar el curso de la conversación. Se trata de un rol normalmente asignado al periodista, aun cuando en ocasiones –y por regla general durante un breve espacio de tiempo– el entrevistado pueda asumir tal papel. La asimetría se manifiesta también en la distinta distribución del poder interaccional, a través de lo que se ha dado en llamar ‘dominancia discursiva’ (Linell, Luckmann 1991; para una ejemplificación de los tipos de dominancia discursiva en *Salvados*, véase Ortega Garrido 2022a). Así, Évole altera el discurrir de la entrevista periodística y la transforma en una conversación espontánea natural. No obstante, a juzgar por los postulados de la teoría periodística contemporánea, se trata de una tendencia bastante extendida en nuestros días. Cantavella Blasco, por ejemplo, concibe sin ambages este género del periodismo oral como una “conversación distendida”, “un diálogo vivo y fresco”, “una conversación que se presupone desenfadada” (2014, pp. 322, 327), en la línea de lo que caracteriza a una interacción de tipo espontáneo.

Otro aspecto que es necesario subrayar atañe a las personas implicadas en el intercambio comunicativo cuando este se emite en directo o bien se distribuye por medios similares, como el *streaming* o la televisión a la carta que ofrecen muchos canales y páginas web en la actualidad. Habría que concebir un espacio múltiple dividido en dos planos: el de la interacción propiamente dicha entre entrevistador y entrevistado, donde se perciben los clásicos papeles de emisor y receptor, y el de la visualización del programa, donde se encuentra el destinatario final de todo lo dicho en el diálogo, es decir, la audiencia. Dicha audiencia, que bien podríamos denominar

‘superreceptor’ o ‘suprarreceptor’, se sitúa por encima de los actores del plano original. La importancia del destinatario radica en que todo el discurso, tanto el del entrevistador como el del entrevistado, se concibe en función de la existencia de tal destinatario –implícita en el momento de la entrevista y proyectada hacia el futuro–, con todo lo que ello conlleva, como señala Escandell (2006, p. 29):

El destinatario es siempre el receptor elegido por el emisor. Pero no sólo eso: además, el mensaje está construido específicamente para él. Este hecho es de capital importancia, ya que condiciona en gran medida la forma del mensaje.

A este destinatario aluden ocasionalmente los participantes en *Salvados* (Ortega Garrido 2021), ya que son conscientes de su presencia y de su papel determinante. El empleo de un tipo de registro concreto, como pueda ser el coloquial, deriva de la importancia del destinatario, a quien los actores involucrados en la entrevista pretenden acercarse a través de esta estrategia lingüística, una de las muchas presentes en el programa (véase Ortega Garrido 2022b).

2. Corpus de estudio y enfoque metodológico

Desde un punto de vista metodológico, partimos de las consideraciones teóricas de la disciplina del análisis del discurso, y más en concreto del análisis de la conversación; por otro lado, nos apoyamos en las herramientas que nos brinda la lingüística de corpus, en tanto en cuanto hacen posible complementar el análisis cualitativo con un análisis cuantitativo, de modo que pueda establecerse con mayor rigor la validez de ciertas observaciones, así como validar o invalidar intuiciones previas.

El corpus de estudio en el que se basa nuestro análisis comprende la transcripción de 23 entrevistas a políticos, en su mayor parte españoles y ocasionalmente hispanoamericanos. En su conjunto, nos acercamos a las 18 horas de emisión del programa, que conforman así un corpus de 148 713 palabras en total. Para afinar el análisis, hemos confeccionado dos subcorpus, denominados ‘Solo Évole’ y ‘Todos menos Évole’, de modo que fuera posible analizar frecuencias de uso presentes sólo en el periodista o sólo en los entrevistados. El análisis lexicométrico lo hemos llevado a cabo con la herramienta informática Sketch Engine, que nos permite, además, la extracción de palabras clave tanto del corpus completo como de los dos subcorpus citados, mediante la confrontación con el corpus de español general actual Spanish Web 2018 (EsTenTen); igualmente, hemos extraído las palabras clave fruto de la comparación de los dos subcorpus entre sí. Analizaremos el léxico mediante la observación de las palabras clave de ambos subcorpus y de sus agrupaciones contextuales más frecuentes.

Por lo que respecta al acercamiento cualitativo, nos detendremos especialmente en el análisis de los usos de ciertos vocablos de marcado carácter coloquial en el español contemporáneo de España que tienen un número de recurrencias relevante en el corpus, como son ‘hostia’, ‘macho’ y ‘joder’.

3. El componente coloquial en la conversación espontánea y en la entrevista periodística

3.1. *Discurso coloquial y conversación espontánea*

Definir las características de la lengua coloquial se revela una tarea compleja, habida cuenta de las dificultades que, de entrada, ya nos encontramos para discernir con claridad lo que se entiende por coloquial (al respecto, véase Briz Gómez 2001, pp. 35-43). En todo caso, siguiendo a Briz Gómez (1996, *passim*), es posible identificar una serie de componentes que se encuentran presentes en el discurso generalmente calificado de tal modo y que, acaso, podrían tomarse como características generales de estos intercambios lingüísticos. Por un lado, para que puedan ser considerados coloquiales deben consistir en conversaciones espontáneas; así, podríamos señalar como consustanciales a este tipo de interacción los catorce puntos señalados por Sacks, Schegloff y Jefferson (1974). Especialmente deben ser tenidos en cuenta los puntos séptimo y octavo, que establecen la espontaneidad de los intercambios a partir de la no especificación de la duración de la conversación (“La duración de una conversación no se especifica previamente”) y de la ausencia de contenidos preparados por parte de los participantes (“Lo que dicen los hablantes no se ha especificado previamente”) (citamos a partir de Tusón Valls 1997, p. 68).

Además, cabría señalar una cuestión referida al contenido, pues en este tipo de discurso prevalecen los temas de la vida cotidiana, generalmente ligados, por añadidura, al momento y al lugar en que se está hablando. Por otro lado, la conversación coloquial espontánea mantiene a lo largo de su desarrollo un carácter informal, entendido este como una liberación, en la medida de lo posible, de cualquier regla que restrinja el fluir del discurso; a este respecto, es de señalar la ausencia de jerarquización entre los interlocutores, con una relación de igualdad entre ellos establecida de manera tácita y que subyace a la conversación de forma implícita y natural. A todo ello se suma, además, la total ausencia de planificación en lo que se dice, como hemos señalado.

Aparte de estas cuestiones de tipo pragmático, existen algunas otras apreciaciones ligadas específicamente a la propia expresión lingüística, sobre todo en lo que se refiere al léxico. En efecto, lo que se entiende por lengua de

registro coloquial se corresponde en general con una frecuencia destacable de expresiones idiomáticas y de un cierto tipo de vocabulario que podríamos considerar excluido de otras modalidades de comunicación o de formas de expresión vehiculadas a través de la lengua escrita, por ejemplo. Así, las palabras consideradas socialmente como malsonantes serían peculiares de una forma de expresión que se asociaría de manera automática con el discurso coloquial. Por otro lado, Briz Gómez (1996, pp. 60-62) señala también que este se caracteriza por una reducción del caudal léxico, como se observa, por ejemplo, en el caso del uso excesivo de los llamados verbos comodín, a saber, ‘hacer’, ‘tener’ o ‘poner’, los cuales se prefieren en este registro frente a las combinaciones más elaboradas que, por su parte, caracterizarían al discurso alejado del registro coloquial. Asimismo, en relación con el léxico, se puede destacar el empleo de un vocabulario específico identificable con el argot; a este respecto, las jergas juveniles serían un buen ejemplo de ello.

Teniendo en cuenta estos factores, cabe preguntarse cuál es el papel que desempeña este nivel específico de lengua en la entrevista periodística.

3.2. *Discurso coloquial y entrevista periodística*

La entrevista periodística se entiende como una transposición al ámbito de la información de lo que sería una conversación donde uno de los participantes recaba datos a partir de otro. Se concibe incluso como un género independiente dentro del periodismo mediante el cual se busca no solo informar, sino también obtener información, conocer las opiniones del entrevistado acerca de ciertos temas e incluso formar culturalmente a la sociedad (Martínez Vallvey 1995, p. 56). La presencia del destinatario, por otra parte, puede formar parte esencial de la definición, como sugiere Cuenca (2013, p. 526):

...la entrevista es un género dialógico, estructurado en turnos, mayoritariamente de pares adyacentes de pregunta breve y respuesta larga, con dos participantes activos y un participante receptivo (el público) que, en principio, presupone un respeto al principio de cortesía, neutralidad y cooperación entre entrevistador y entrevistado.

El destinatario no solamente es tenido en cuenta en el momento de la elocución, sino que existe una retroalimentación según la cual la lengua del espectador se ve influenciada por la lengua de los medios de comunicación, pues, de entre las distintas variedades de lengua, “no hay duda de que es la periodística la que más influye en el hablante común; en nuestro caso, en la actuación lingüística de los hablantes españoles” (Romero Gualda 2008, pp. 9-10). De ese modo, se crea un círculo vicioso donde la lengua coloquial de los espectadores se emplea como recurso por parte de periodistas y

entrevistados, lengua coloquial que a su vez incide en la forma de hablar de los espectadores. En efecto, hace décadas que se señala ese efecto de relajación en cuanto a propiedad y corrección lingüística, por lo que tampoco es de extrañar la presencia de palabras y expresiones vulgares, barbarismos y errores varios en los distintos espacios televisivos. Todo ello refleja una “reducción de la complejidad”, estrategia encaminada a captar la atención del público y acercarse a él, como apuntaba Arfuch (1995, p. 117):

Si bien cada modalidad de entrevista tiene sus reglas, lo que puede notarse en general es una flexibilización de la palabra, la posibilidad de usar imágenes familiares, irónicas, vulgares, el recurso al chiste o al refrán. En ciertos casos, sobre todo cuando se acentúan aspectos técnicos (esa jerga de especialistas, especialmente en economía, que cada vez más parece necesitar de traducción), el discurso político en la entrevista opera, al estilo de la divulgación científica, por reducción de la complejidad.

Igualmente, Briz Gómez señalaba por las mismas fechas un fenómeno similar, es decir, el “proceso de coloquialización creciente en TV y radio, favorecido, entre otras razones, por la tendencia a la comunicación interactiva y por la constante lucha por la audiencia” (1996, p. 11). En los debates televisivos y las tertulias políticas radiofónicas de contenido político, tan de moda en España durante los últimos lustros, se da un paso más, pasando de lo coloquial a lo descortés, especialmente mediante los “colaboradores no expertos [...], que a través de la coloquialidad y la descortesía «imitan» una conversación distendida” (Fuentes Rodríguez 2013, p. 20).

Por lo que respecta a Évole, ya hemos señalado su intención (consciente o no) de renovar el género periodístico de la entrevista, cosa que lleva a cabo valiéndose de un estilo personal y de una puesta en escena inusitada. Ese estilo característico encuentra en la lengua coloquial uno de sus rasgos principales, dotando así al encuentro comunicativo de una peculiar y novedosa frescura a los ojos de los telespectadores, atraídos en parte por ese registro cercano a ellos que roza con la espontaneidad. Bien es cierto que se mantienen en buena medida determinadas reglas del género, de modo que podemos seguir llamándolo entrevista, pero, como decimos, con una serie de elementos novedosos. La estructura de preguntas y respuestas deja paso a una interacción donde domina o intenta dominar la espontaneidad. Para ello, en contra de la tradición del género, que exigía del periodista un carácter serio, Évole parte del principio de la coloquialidad del encuentro como base para su renovación, ayudado por la frecuente utilización de vocablos y expresiones pertenecientes a ese registro, como tendremos oportunidad de observar en los siguientes apartados. Por su parte, los políticos entrevistados no suelen recurrir con tanta insistencia al léxico coloquial, sino que, en líneas generales, se mantienen en un terreno de mayor formalidad, aunque a veces se dejan llevar por el tono del periodista, teñido de espontaneidad. Por otro lado, se

valen de expresiones idiomáticas que tienen un valor metafórico, como por ejemplo ‘poner la mano en el fuego’ (empleado sobre todo por Esperanza Aguirre y Jorge Fernández Díaz en sus respectivas entrevistas), con una precisa función pragmática de acercamiento al telespectador:

En general se puede decir que el uso de la mayor parte de estas fórmulas, por su carácter coloquial, da un tono distendido, menos formal, que aproxima el discurso al hombre de la calle, al tiempo que facilita la rápida comprensión del argumento; de ahí la eficacia pragmática de estas expresiones fijas. (Fernández Lagunilla 2009, p. 72)

4. Palabras clave en el corpus de estudio

4.1. Palabras clave en cada uno de los subcorpus

Como hemos señalado anteriormente, la extracción de palabras clave, herramienta que se ha demostrado fundamental cuando nos valemos de la metodología de la lingüística de corpus (Scott 1997; Baker 2006), nos da la oportunidad de comprobar cuál es la tematicidad o *aboutness* del texto (véase Phillips 1989, pp. 7-10, 26, 53-54, citado en Gabrielatos 2018, p. 225). Por un lado, hemos extraído las palabras clave de cada uno de los subcorpus comparándolos entre sí. Así, en el subcorpus ‘Solo Évole’ podemos observar que entre las primeras palabras clave, aparte de varios nombres propios, figura en séptima posición el sustantivo ‘tío’ (Tabla 1), que corresponde al sentido coloquial que recoge el *Diccionario de la lengua española* de la Real Academia Española (en adelante *DLE*) en sus acepciones sexta (“Persona de quien se pondera algo bueno o malo”) y séptima (“Persona cuyo nombre y condición se ignoran o no se quieren decir”), como podemos observar a través de las concordancias de nuestro corpus, ejemplos de (1) a (5):

- (1) Con el tío este acabó muy mal usted.
- (2) Pero es que el otro no era un tío cualquiera, era..., era un señor que...
- (3) No sé, yo digo, hostia, este tío o tiene mucha moral o es que realmente...
- (4) Cuéntame, tío.
- (5) ...a usted la nariz le ha dicho: uf, este tío no está aquí porque le interese la política, este tío está aquí para enriquecerse.

	Palabras clave	Frecuencia absoluta	Frecuencia por millón
1	urdangarín	17	273,88
2	iñaki	17	273,88
3	entrevistar	13	209,44
4	miguel	12	193,33
5	marcial	10	161,11
6	merkel	7	112,78
7	tío	6	96,67
8	caixa	5	80,55
9	salpicar	5	80,55
10	caracas	5	80,55

Tabla 1

Diez primeras palabras clave del subcorpus ‘Solo Évole’ frente a ‘Todos menos Évole’.

Por otro lado, el verbo ‘salpicar’, que ocupa la novena posición en la tabla, siempre se emplea en las entrevistas analizadas para hablar de casos de corrupción que salpican a los políticos, en una construcción que colorea la lengua del periodista con una expresión idiomática de valor metafórico similar a las que señalábamos en las últimas líneas del epígrafe anterior.

Por su parte, en el caso de los políticos entrevistados, las palabras clave en comparación con el discurso del entrevistador revelan un grado elevado de formalidad, con sustantivos y verbos que se refieren a conceptos abstractos y con adverbios en -mente (Tabla 2).

	Palabras clave	Frecuencia absoluta	Frecuencia por millón
1	distinto	43	409,73
2	resolver	34	323,97
3	absolutamente	34	323,97
4	asunto	31	295,38
5	efectivamente	30	285,86
6	etcétera	28	266,80
7	determinar	25	238,21
8	función	24	228,68
9	posición	20	190,57
10	nación	19	181,04

Tabla 2

Diez primeras palabras clave del subcorpus ‘Todos menos Évole’ frente a ‘Solo Évole’.

Es de señalar que un vocablo como ‘nación’ pertenece a las llamadas ‘palabras-símbolo’ o ‘palabras-emblema’, carentes de determinación y, por ello mismo, situadas en el ámbito de la abstracción (Fernández Lagunilla 2014, p. 22).

Por otra parte, podemos extraer las palabras clave de ambos subcorpus en relación con un corpus general de español de grandes dimensiones. En nuestro caso nos valemos de Spanish Web 2018 (esTenTen), un corpus general de español actual con miles de millones de palabras. De la misma manera que sucedía con las palabras clave vistas anteriormente, en el caso del periodista se observan vocablos adscribibles al ámbito coloquial ya en las diez primeras posiciones, en concreto ‘hostia’, que ocupa nada menos que el primer lugar, y la interjección ‘ah’ (Tabla 3).

	Palabra clave	Frecuencia absoluta	Frecuencia por millón
1	hostia	45	724,98
2	urdangarín	17	273,88
3	president	18	289,99
4	otegi	9	144,99
5	bonsáis	6	96,66
6	ah	90	1449,97
7	peares	5	80,55
8	pdcát	5	80,55
9	rubalcaba	17	273,88
10	junqueras	11	177,21

Tabla 3

Diez primeras palabras clave del subcorpus ‘Solo Évole’ frente a Spanish Web 2018.

En cambio, en el caso del subcorpus ‘Todos menos Évole’ (Tabla 4), los únicos resultados que podrían considerarse coloquiales son las interjecciones ‘eh’, en cuarta posición, y ‘ah’, que figura en duodécima posición.

	Palabra clave	Frecuencia absoluta	Frecuencia por millón
1	aló	11	104,81
2	espe	8	76,22
3	revisable	7	66,69
4	eh	60	571,71
5	esquerra	11	104,81
6	jordi	47	447,84
7	referéndum	52	495,48
8	jaboís	4	38,11
9	símil	9	85,75
10	croissant	5	47,64

Tabla 4

Diez primeras palabras clave del subcorpus ‘Todos menos Évole’ frente a Spanish Web 2018.

En los siguientes epígrafes nos detendremos en el análisis de las combinaciones contextuales en que se encuentran las palabras clave pertenecientes de manera más clara al ámbito coloquial, a saber, ‘hostia’ y ‘tío’.

4.2. Contextos de uso de las palabras clave ‘hostia’ y ‘tío’

Hemos visto en el apartado anterior que ‘hostia’ resulta ser la primera palabra clave en el discurso de Évole cuando se compara con un corpus general de español contemporáneo. En total, se repite hasta en 45 ocasiones a lo largo de las veintitrés entrevistas que componen nuestro corpus. Por el contrario, en el discurso de los políticos entrevistados solamente encontramos tal palabra en boca de Pablo Iglesias, el exdirigente de Podemos, cuyo modo de hablar, de la misma manera que sucede con el periodista, quiere acercarse a la forma de expresión más popular y coloquial. El por entonces líder de Podemos llega a pronunciar hasta en cuatro ocasiones tal palabra, como se puede comprobar a través de las concordancias:

- (6) ...cuando ves cómo mejora el nivel de vida, cómo en Ecuador, por ejemplo, hostia, invierten en innovación, tienen doctores españoles trabajando en centros de excelencia...
- (7) Hostia, es que eso molaría, es que eso molaría, es decir, dar la cara, es decir, no sé... Como un presidente del Gobierno que se junta cada mes con cinco periodistas así cabrones como tú, que en lugar del rollo de ahora de pantalla de plasma, rueda de prensa con dos preguntas de los colegas, hostia, igual sí.
- (8) No, hostia, lo que escribimos, no, sino que me veo dirigiendo un Consejo de Ministros y diciendo “esto lo tenemos que hacer”.

Además, aparte de ‘hostia’, detectamos en su discurso otras voces propias del léxico coloquial, especialmente en el ejemplo (7), donde se acumulan palabras y expresiones idiomáticas características de este registro, como “molaría”, “dar la cara”, “cabrones”, “rollo”, “colegas” o “igual”.

En boca del entrevistador la situación es similar, en el sentido de que a la voz ‘hostia’ a veces la acompaña algún otro vocablo coloquial, como por ejemplo la otra palabra clave que hemos mencionado, es decir, ‘tío’, así como algunas más pertenecientes a ese mismo nivel de lengua (“rojillo”, “fotón”, “flipo”), en ocasiones formando parte de expresiones idiomáticas (“Hostia, Pedro, nos dejas colgados”, en la entrevista a Pedro Sánchez; “Hostia, me está pegando aquí un chorreo, presidente, impresionante ahora”, en la entrevista al presidente del Gobierno Mariano Rajoy, donde el periodista no duda en introducir el vulgarismo varias veces, a pesar del cargo político desempeñado por el entrevistado).

Por lo que respecta a ‘hostia’, podemos comprobar su diferente distribución sintáctica en estas entrevistas: el uso más frecuente corresponde al de la mera exclamación únicamente con esa palabra, clasificada por el *DLE*

como interjección malsonante utilizada en España que denota “sorpresa, asombro, admiración, etc.” Es el caso de los ejemplos (9) y (10), que corresponden a una entrevista al líder nacionalista catalán Artur Mas, así como de (11) y (12), en entrevistas a los políticos del Partido Popular Esperanza Aguirre y Alberto Núñez Feijoo, respectivamente (señalamos siempre en cursiva la frase donde se encuentra la palabra o expresión coloquial en cuestión):

- (9) Artur Mas: Si esto sale bien es mérito de todos y si no sale bien es culpa mía.
 Jordi Évole: *Hostia, me recuerda usted a un entrenador de fútbol.*
 Artur Mas: ¿Verdad?
 Jordi Évole: Que no es el equipo quien pierde... la responsabilidad es mía.
- (10) Artur Mas: Esto es un ascensor...
 Jordi Évole: ...está muy bien camuflado.
 Artur Mas: ...del siglo XXI.
 Jordi Évole: Pero está... *Hostia, es que cuando lo ves...*
 Artur Mas: Sí, sí, parece...
 Jordi Évole: No parece un ascensor, pero lo parece. Por aquí parece un pasadizo secreto.
 Artur Mas: Pues no, lo es.
- (11) Jordi Évole: Estaba leyendo aquí lo del cumpleaños de Antena 3.
 Esperanza Aguirre: Sí, veinticinco años de Antena 3.
 Jordi Évole: Estuviste.
 Esperanza Aguirre: Estuve, estuve, estuve.
 Jordi Évole: *Hostia, no te pierdes una.*
 Esperanza Aguirre: Me pinté las uñas de verde, porque iba a ir de verde.
 Jordi Évole: Por La Sexta.
 Esperanza Aguirre: Sí. No, no, no, no, no, eso... No había caído en eso. Porque llevaba un vestido verde a este acto. Esto sí.
- (12) Jordi Évole: Pero a mí me han dicho que... que usted era como... como rojillo, de izquierdas.
 Alberto Núñez Feijoo: Yo voté al Partido Socialista, sí,
 Jordi Évole: ¿En el año 82? ¿A Felipe González?
 Alberto Núñez Feijoo: En el año 82.
 Jordi Évole: ¿Solo en esa ocasión?
 Alberto Núñez Feijoo: Pues creo que no sé si le voté en el 86 también, eso yo creo que sí, claro.
 Jordi Évole: *Hostia, pero eso ya es reincidente.*
 Alberto Núñez Feijoo: Sí, sí, sí, claro, si yo creo que Felipe era un buen presidente.

En ocasiones, Évole recurre a una estrategia que consiste en introducir en estilo directo palabras que podría haber pronunciado él mismo u otra persona, a modo de cita no auténtica pero sí verosímil. Este recurso, además, permite al periodista expresar lo que se consideraría una verdad asumida por la mayor parte de la comunidad. En el corpus analizado, se observa con frecuencia que estas frases en estilo directo vienen introducidas precisamente por la voz coloquial ‘hostia’. Así sucede en el ejemplo (13), perteneciente a la entrevista

al secretario general del PSOE Pedro Sánchez, donde observamos que se recurre al estilo directo para rebajar la fuerza del juicio expresado en la hipotética frase, atribuyéndola a una voz ausente, la voz del propio entrevistador reflexionando consigo mismo. Esta reflexión, en todo caso, se abre con la interjección de asombro, reforzada también por un verbo coloquial que expresa sorpresa empleado poco antes (“yo flipo”):

- (13) Jordi Évole: Yo flipo cuando ayer te escucho decir: “voy a coger el coche”, “voy a recorrer otra vez España”... No sé, yo digo: “*hostia, este tío o tiene mucha moral o es que realmente lo del poder llama mucho*”.
Pedro Sánchez: No, yo es que creo en la política.

De igual modo sucede en (14), durante la entrevista a la expresidenta de la Comunidad de Madrid Esperanza Aguirre, en la cual el periodista expone su sorpresa primero de manera explícita (“a mí me llama la atención”) y a continuación mediante una frase en estilo directo que recogería sus propios pensamientos, de nuevo introducida por “hostia”:

- (14) Jordi Évole: ¿Bárceñas les ha engañado a todos?
Esperanza Aguirre: Yo es que a Bárceñas tengo que confesarle que es que yo estuve siempre aquí y no he tenido relación con él. Nada. Cero.
Jordi Évole: ¿Ni una llamada?
Esperanza Aguirre: No.
Jordi Évole: ¿Ni un día por teléfono?
Esperanza Aguirre: Cero. Pues fíjese que no.
Jordi Évole: A mí me llama la atención, porque yo me decía: “*hostia, gente muy preparada, que alguien como Bárceñas haya podido engañar a Aznar, Rajoy, Arenas, a Cascos, a Rato, a usted*”.
Esperanza Aguirre: Vamos a ver. Lo importante de Bárceñas es que se llegue al conocimiento de la verdad.

La oración en estilo directo también puede corresponder a unas hipotéticas palabras pronunciadas o imaginadas por el entrevistado, como en el ejemplo (15), en que el periodista atribuye a Núñez Feijoo un pensamiento determinado con la finalidad de incitar al político a responder y explicar su versión de los hechos. En este caso, además de “hostia”, la oración incluye el aumentativo con valor coloquial “fotón”:

- (15) Alberto Núñez Feijoo: Yo no voy a un sitio para coger una manguera, sino que estábamos haciendo un directo, yo creo, y hubo un rebrote de detrás de un incendio y una persona me dio la manguera y yo la cogí, vamos, eso es una... es una ingenuidad bastante acusada.
Jordi Évole: Pero usted coge la manguera y dice: “*hostia, qué fotón va a salir mañana en la portada de La Voz de Galicia*”.
Alberto Núñez Feijoo: Mire usted, yo creo que ya la he calificado. No.

En otros casos, esa voz en estilo directo representaría la opinión generalmente

aceptada, lo que se considera *vox populi*. De esta manera, el periodista se distancia de lo que está diciendo, ya que atribuye sus palabras a la gente, más que a sí mismo, salvaguardando de este modo su imagen positiva. Así ocurre en dos momentos de la entrevista a Rajoy; en (16), el entrevistador introduce un parlamento atribuido a un supuesto amigo suyo que se abre con la interjección malsonante:

- (16) Jordi Évole: Hay un amigo mío que me ha comentado: “*hostia, qué mal lo tiene que ver Rajoy para que os dé a vosotros una entrevista*”.
 Mariano Rajoy: ¿Ah, sí? ¿Y quién ha sido ese amigo?
 Jordi Évole: Un amigo de Cornellá.

Más adelante, ejemplo (17), se vuelve a introducir un hipotético discurso directo del amigo, al cual se le añade otro más en boca de la gente en general, que incluye “hostia” como interjección:

- (17) Jordi Évole: Mi amigo, el del principio, el que me ha dicho...
 Mariano Rajoy: El de Cornellá.
 Jordi Évole: Sí, el de Cornellá.
 Mariano Rajoy: Me lo tiene que presentar.
 Jordi Évole: Sí, hombre, algún día se lo presentaré. Un día se viene a Cornellá y se lo presento. También me ha dicho, dice: “yo creo que este tipo de frases a Rajoy le están humanizando”. La gente pues dice: “*pues, hostia, yo también me equivoco*”.
 Mariano Rajoy: Le voy a decir una cosa: es que los políticos también somos seres humanos y tenemos sentimientos.
 Jordi Évole: Esa es otra frase que también ha hecho mucha fortuna, pero al revés.

El estilo directo de contenido hipotético aplicado a una tercera persona puede igualmente referirse a compañeros de partido del político entrevistado, como en el ejemplo (18), perteneciente a la entrevista a Pedro Sánchez, donde, además, se recurre a la expresión coloquial ‘dejar colgado a alguien’, ya señalada:

- (18) Jordi Évole: Ayer dijiste que no le querías fallar a la militancia y tampoco le querías fallar al partido. ¿Dejando el acta le fallaste a tus compañeros que votaron no a Rajoy?
 Pedro Sánchez: No, porque creo que hablé mucho con ellos.
 Jordi Évole: ¿Y ninguno te ha reprochado: “*hostia, Pedro, nos dejas colgados*”?
 Pedro Sánchez: No, no, no, no, al contrario, no me decían eso.

Por otra parte, en el conjunto de entrevistas analizadas podemos encontrar una serie de expresiones que incluyen la voz ‘hostia’; mediante la herramienta Word Sketch de Sketch Engine, observamos que las combinaciones sintácticas de esta palabra son tres: ‘inflar a hostias’ (“resulta

que un policía me infló a hostias”, en la entrevista al ministro del Interior Jorge Fernández Díaz), ‘dar hostias’ (“creo que te veo preocupado pero por una cosa, porque sin Zapatero, ¿a quién le vais a dar las hostias?”), en una de las entrevistas a Miguel Ángel Rodríguez) y ‘caer hostias’ (“os hubiesen caído hostias como panes”, en la entrevista a Pablo Iglesias).

Por lo que respecta a la voz ‘tío’, la única combinación sintáctica destacable es “este tío”, empleada dentro de frases en estilo directo en boca de Évole, como sucede en las entrevistas a Artur Mas y a Pedro Sánchez, en los ejemplos ya vistos (5) y (13), respectivamente.

5. Otros vocablos coloquiales frecuentes

Dentro de los vocablos coloquiales más frecuentes en el corpus de estudio, aparte de ‘tío’ y ‘hostia’, tenemos alguno más; uno de ellos considerado también malsonante, a saber, ‘joder’, y otro, usado como marcador del discurso, ‘macho’, que se revela frecuente solamente en boca del entrevistador en los primeros años del programa. En efecto, el empleo de ‘macho’ se circunscribe a las tres entrevistas cortas realizadas en 2009 a quien fuera portavoz del Gobierno con el Partido Popular entre 1995 y 1998, Miguel Ángel Rodríguez. En estas conversaciones, el componente coloquial tiene un peso importante, entre otras cosas por el constante tono fuertemente irónico del periodista. Así, la voz ‘macho’ a veces se presenta en combinación con vocablos vulgares, especialmente con ‘hostia’ en (19), (22), (23) y (24) y con ‘joder’ en (19) y (20), así como con ‘putada’ en (21):

- (19) Jordi Évole: Uy, qué azulito todo.
 Miguel Ángel Rodríguez: Sí, sí.
 Jordi Évole: Está a tu gusto, ¿eh?
 Miguel Ángel Rodríguez: Este colegio está instruyendo bien a los chavales.
 Jordi Évole: *Hostia, macho, hostia, vaya interventoras que tiene el PP.*
 Miguel Ángel Rodríguez: Mira cómo son las apoderadas del PP.
 Jordi Évole: *Joder, macho, cómo te has puesto. ¿Pero tú dónde venías? ¿Qué venías a intervenir?*
- (20) Miguel Ángel Rodríguez: Yo te digo lo que hay. Si el PSOE queda por debajo de 110 escaños, yo creo que va a tener que volver Felipe González. No hay nadie en este minuto en el Partido Socialista que pueda rehacer.
 Jordi Évole: *Joder, macho, Felipe González, pero eso era como pedir a Aznar cuando Rajoy estaba en horas bajas.* Ay, perdona, si es que lo pidieron.
 Miguel Ángel Rodríguez: Claro. Vamos a ver cómo se comporta la izquierda sin ningún poder.
- (21) Jordi Évole: No, no, no voy a recuperar cortes tuyos.
 Miguel Ángel Rodríguez: Bueno.
 Jordi Évole: *No te voy a hacer esa putada. Porque, macho, les has dado por todos lados.*
 Miguel Ángel Rodríguez: No, lo que pasa es que me he defendido de los

ataques que han tenido, tan sectarios...

Jordi Évole: Ah, vale, vale, que son defensas.

Destaca sobre todo una frase pronunciada por el entrevistador en el ejemplo (22), formada exclusivamente por tres de las palabras que estamos comentando, propias de un registro coloquial (“Hostia, macho, tío”):

- (22) Jordi Évole: He preparado unos eslóganes.
 Miguel Ángel Rodríguez: A ver.
 Jordi Évole: Mariano decente, Mariano presidente.
 Miguel Ángel Rodríguez: Sí, no está mal. A ver.
 Jordi Évole: Vota Mariano. Se hace los trajes con la mano.
 Miguel Ángel Rodríguez: Ja, se hace los... A mano. Mira, Mariano no es de sastre. Y ahí tenemos, ahí tenemos un... eslogan que podemos pillar.
 Jordi Évole: *Hostia, macho, tío*. Mira que yo esto me lo he estao..., me lo he estao...

Asimismo, nótese la pronunciación relajada de algunas palabras, con desaparición del alófono consonántico dental fricativo sonoro, específica de la lengua oral y más concretamente del nivel coloquial: “estao” (22), “decantao” (23) y “joer” (24). Debe tenerse en cuenta que no se produce tal pérdida en todas las ocasiones en que aparece un participio de la primera conjugación o cualquier otra palabra susceptible de perder la -d-intervocálica, pues los niveles de habla son múltiples. En efecto, el propio periodista unas veces mantiene una pronunciación estándar con la dental intervocálica presente y otras pasa a un registro de habla más popular con la consecuente eliminación de la consonante, y del mismo modo sucede con los entrevistados; cuando la eliminación de la consonante es patente, así lo reflejamos en la transcripción.

- (23) Jordi Évole: Hostia, pues qué matrimonio más divertido, ¿no?
 Miguel Ángel Rodríguez: Bueno, y con mis hijos ni te cuento.
 Jordi Évole: ¿Sí?
 Miguel Ángel Rodríguez: Sí.
 Jordi Évole: ¿Qué? ¿Te han salido del PSOE?
 Miguel Ángel Rodríguez: Una del PSOE y el otro todavía no está muy decantao, pero... pero, ahí estoy en la pelea.
 Jordi Évole: *Pero, macho, eso es como si eres del Madrid y se te hace del Barça*.
 Miguel Ángel Rodríguez: Es que además es del Barça.
 Jordi Évole: *Hostia, macho, pero tú tienes un cuadro clínico en tu casa*.
 Miguel Ángel Rodríguez: Nos queremos.
- (24) Miguel Ángel Rodríguez: Lo que hay que hacer para salir por la tele, tío.
 Jordi Évole: Ven, Miguel Ángel.
 Miguel Ángel Rodríguez: Oye, pero esto no será un fusilamiento o algo, ¿no?
 Jordi Évole: No, no, no. Esa época se ha acabado. Cuidado. Ten cuidado.
 Miguel Ángel Rodríguez: Pero, avisad, joer.
 Jordi Évole: *Hostia, macho, estás ciego*.

Miguel Ángel Rodríguez: Es que yo no me conozco además la furgoneta.

Por su parte, la interjección coloquial ‘joder’, señalada igualmente como malsonante por el *DLE*, cuenta con 15 recurrencias en el corpus. Esta palabra, exclusiva del léxico coloquial, refuerza ese contacto entre el entrevistador y la lengua de corte más popular hablada en España. En las entrevistas analizadas, el uso de ‘joder’ como verbo se vehicula a través de las siguientes combinaciones: “no es por joder”, “te vas a joder”, “nos ha jodido”, “es jodido” y “que se jodan”.

De todos modos, este vocablo se emplea más frecuentemente como interjección, ejemplos (25), (26), (27) y (28), en una ocasión seguida de preposición (“joder con”), en (29):

- (25) Jordi Évole: Perdona, señor Feijoo, no sé si se está emocionando o es que le está... Le cae una lágrima de...
 Alberto Núñez Feijoo: Es que hace 3 grados aquí.
 Jordi Évole: Ah, vale, vale.
 Alberto Núñez Feijoo: A pesar de que no lo note.
 Jordi Évole: *Joder, digo, se está emocionando hablando de ser de izquierdas, tener corazón...*
- (26) Jordi Évole: De presentarse en todos los municipios a no presentarse a ninguno con la marca Podemos, *da la sensación de, joder, que habéis tenido miedo.*
 Pablo Iglesias: Hemos sido prudentes porque en política hay que ser prudentes y no hay que ser arrogantes y hay escenarios que te permiten ganar.
- (27) Jordi Évole: Ostras, qué maravilla, qué cantidad de luces. Esto parece un plató de televisión.
 Josep Antoni Duran i Lleida: Sí.
 Jordi Évole: *Joder, qué hemiciclo.*
- (28) Jordi Évole: Y con esos nueve puntos, ¿cuánto rato puede llenar de mitin?
 Alfredo Pérez Rubalcaba: Depende, igual alguno ni siquiera le toco. Tienes la idea de lo que tratas de decir ese día, fundamentalmente lo piensas en función de pues eso es lo que hay.
 Jordi Évole: *Joder, qué tiempo.*
 Alfredo Pérez Rubalcaba: ¿Vamos para allá?
- (29) Jordi Évole: ¿Ha caído usted en muchas contradicciones?
 José Mujica: No, en lo fundamental en ninguna, pero a veces he tenido que darle un sablazo a un compañero, bueno, útil.
 Jordi Évole: ¿Un sablazo qué es?
 José Mujica: Un sablazo es echarlos y... queriéndolos y... valorándolos. Les tuve que apretar el pescuezo. Eso es muy doloroso, cuando tú tienes que encuadrar a un hijo tuyo y hasta cometer alguna injusticia con él porque al todo general le es conveniente. Eso es doloroso.
 Jordi Évole: *Joder con la política.*
 José Mujica: Sí.

En el ejemplo (30), al igual que con ‘hostia’, la interjección ‘joder’ abre un discurso en estilo directo atribuido a una persona indeterminada:

- (30) Jordi Évole: Puede haber gente que te esté viendo y esté pensando: “*joder, esto lo crítica mucho, pero bien que se lleva sus 5000 euros al mes*”.
 José Manuel Pérez Bouza: Yo creo que una cosa no tiene nada que ver con la otra.

Esta palabra la emplea principalmente Évole, aunque la podemos encontrar en boca de dos de los entrevistados. Uno de ellos es Pablo Iglesias, quien intenta acercarse a una forma de expresión popular, como ya hemos visto también en el caso de ‘hostia’. En (31), el dirigente de Podemos usa ‘joder’ como interjección, mientras que en (32) aparece como forma verbal:

- (31) Jordi Évole: ¿Crees que en países como Ecuador, donde Correa lleva gobernando desde hace siete años, o en Venezuela, donde el partido de Chávez, ahora con Maduro, lleva gobernando desde 1999, también hay casta?
 Pablo Iglesias: Seguro que sí, seguro que sí y eso es...
 Jordi Évole: ¿Por qué nunca habláis de esa casta?
 Pablo Iglesias: *Joder, claro que hablamos.*
 Jordi Évole: ¿Sí?
 Pablo Iglesias: Si tú vieras las cosas que...
 (32) Pablo Iglesias: Es una cosa que hacen los suizos, que, fíjate tú que en Suiza, que tienen...
 Jordi Évole: ¿Ah, sí? ¿Lo hacen los suizos?
 Pablo Iglesias: Eso, creo que son...
 Jordi Évole: ¿Qué hacen?
 Pablo Iglesias: ...no sé si 20 veces como máximo el salario mínimo, una cosa así. *Esto no es por joder*, no es que a usted le estemos limitando el salario máximo para putearle, lo que estamos diciendo es: “para que el país vaya mejor hay que evitar la desigualdad, que es lo que destruye las economías de los países”.

El otro político es el senador gallego José Manuel Pérez Bouza, el cual igualmente emplea en (33) tanto la interjección como la forma verbal:

- (33) José Manuel Pérez Bouza: Aquí se trabaja, aquí se hacen cosas importantes. Aquí no se hace nada. *Joder, nada. Nada.* Lo que hacemos es el tonto. Yo a veces que vienen niños y que viene público y tal, yo... Qué imagen se llevan. Y la mayoría de la gente no sabe exactamente que el Senado no sirve para nada, porque, si no, pues habría más presión y efectivamente tendrían que hacer algo PP y PSOE y siempre se nos ponía: “no, es que hay que reformar la Constitución, es que eso es sagrado, es que hace falta consenso” y tal. *Y luego, joder, en dos tardes reforman la Constitución y se quedan tan panchos.* Puestos a reformar, eliminemos las diputaciones, reformemos el Senado. “No, no, eso no toca”. *Ahora lo que toca es satisfacer a los mercados para que nos sigan jodiendo.*

Por otro lado, aparte de la palabra en cuestión, en los ejemplos anteriores es posible identificar otras expresiones del registro coloquial, como “dar un

sablazo”, “apretar el pescuezo” (29), “putearle” (32), “lo que hacemos es el tonto”, “y tal” o “y se quedan tan panchos” (33).

El empleo de estas palabras malsonantes, como vemos, se caracteriza por la combinación con algunas otras de semántica cercana (como, precisamente, en el caso de ‘putear’), de modo que el diálogo alcance esa caracterización de conversación espontánea, salpicada de términos coloquiales, tal como sucede en el habla corriente de España. Por otro lado, el uso por parte de Évole de este recurso lingüístico parece extrapolarse en ocasiones al político entrevistado, como en algunos de los casos ya vistos o en (34), pasaje de la entrevista a Esperanza Aguirre donde la expresidenta de la Comunidad de Madrid emplea al final el tono coloquial (“eres un cachondo, macho”) como maniobra de distracción ante un ataque implícito a la alcaldesa de Madrid Ana Botella y a su discurso en inglés de septiembre de 2013 para apoyar la candidatura de la capital española a los Juegos Olímpicos de 2020:

- (34) Jordi Évole: Yo no sé tanto inglés como usted, que usted además...
 Esperanza Aguirre: No te creas que sé mucho, pero en fin.
 Jordi Évole: Me han mandado un ejercicio para este fin de semana, que estaba aprendiendo, y tengo que traducir la siguiente frase: “tomarse una relajante taza de café con leche en la Plaza Mayor”.
 Esperanza Aguirre: Tú eres un cachondo, macho, ¿no? *Relaxing cup of café con leche.*

6. Expresiones coloquiales en forma de glosa: ‘o sea’ e ‘igual’

Otras palabras del léxico coloquial que podemos encontrar con una cierta sistematicidad en nuestro corpus de estudio son el marcador discursivo ‘o sea’ y el adverbio ‘igual’. En el caso de ‘o sea’, el periodista recurre a él para la introducción de glosas, mecanismo mediante el cual se reformula lo dicho en el turno anterior por el otro interviniente (Garfinkel y Sacks 1970; Orletti 2000, pp. 45-71), a veces con una finalidad específica. En efecto, el uso de la reformulación hace posible que un interlocutor cambie el sentido de lo dicho previamente, que simplifique, adapte o, en definitiva, altere las palabras y la intención del otro. Si realizamos un recuento estadístico de la frecuencia de uso de este marcador, podemos observar que su empleo por parte del entrevistador es muy elevado, con 74 recurrencias en todo el corpus, frente a las 39 recurrencias del conjunto de todos los entrevistados (Tabla 5).

	Frecuencia absoluta	Frecuencia por millón
Solo Évole	74	1192,20
Todos menos Évole	39	371,61

Tabla 5
Frecuencia del marcador coloquial ‘o sea’ en los dos subcorpus.

Si nos detenemos en los ejemplos concretos, se revela ese uso pragmático por parte de Évole, que quiere quedarse con la última palabra mediante la reformulación de lo dicho inmediatamente antes por el entrevistado, el cual puede aceptar o no tal reformulación. Por ejemplo, en (35), dentro de la entrevista a Carles Puigdemont, vemos que, tras una glosa interpretativa, el político catalán acepta con pocas reservas la versión del periodista:

- (35) Jordi Évole: Si el día uno se celebra el referéndum con esas garantías que usted dice que se celebrará y sale que sí, porque yo sigo dudando mucho de que pueda salir que no, ¿usted proclama la independencia de Cataluña al cabo de cuarenta y ocho horas?
Carles Puigdemont: Bueno, proclama el Parlamento catalán, y evidentemente la soberanía.
Jordi Évole: *O sea, el miércoles cuatro de octubre el Parlamento hace una declaración de independencia.*
Carles Puigdemont: Sí, sí, podría ser, podría ser, podría ser.

Y otro tanto ocurre en (36), fragmento perteneciente a la entrevista a Esperanza Aguirre:

- (36) Jordi Évole: ¿Por qué el gobierno del Partido Popular no aprobó la dación en pago?
Esperanza Aguirre: Pues no lo sé, porque yo es que estoy absolutamente a favor de la dación en pago.
Jordi Évole: *O sea, ¿le hubiese gustado que su partido hubiese votado a favor de la dación en pago en el Congreso?*
Esperanza Aguirre: Hombre, hombre, pero no le quepa a usted la menor duda.

Por su parte, el adverbio ‘igual’, con un sentido de posibilidad similar al de ‘tal vez’ o ‘quizá’, es considerado en la actualidad voz coloquial por el *DLE* (12^a acepción de la palabra) y por la *Nueva gramática de la lengua española* de la Real Academia (I, p. 1957, punto 25.14ñ), así como por ciertos gramáticos, que también lo sitúan en el registro informal de la lengua (Matte Bon 2011, p. 258). Del análisis de frecuencias de uso se desprende que el periodista cuadruplica el empleo de este adverbio (104 recurrencias) respecto a los políticos entrevistados (25 recurrencias). A modo de ejemplo de este uso coloquial del adverbio ‘igual’, en (37), (38) y (39) reproducimos algunos

momentos de las entrevistas a Jorge Fernández Díaz, Artur Mas y Esperanza Aguirre, respectivamente, señalando en cursiva las frases donde se encuentra tal adverbio:

- (37) Jorge Fernández Díaz: Es que la cadena perpetua es, sería inconstitucional, es que aquí estamos hablando de prisión permanente revisable.
 Jordi Évole: *Bueno, igual como es constitucional le han puesto ese nombre.*
 Jorge Fernández Díaz: No, no, no, es que la prisión permanente revisable...
 [...]

 Jordi Évole: Pero para que yo lo entienda, señor ministro...
 Jorge Fernández Díaz: No, se lo voy a explicar.
 Jordi Évole: Yo, yo le hago la pregunta, *porque igual así me queda más claro.*
 ¿La prisión permanente revisable puede provocar que un terrorista con delitos de sangre que hoy entra en la cárcel no salga hasta que se muera?
- (38) Jordi Évole: Bueno, hoy es 9-N, ya han pasado las diez de la mañana y usted no está detenido todavía.
 Artur Mas: Ja, ja, ja. Ah, bueno.
 Jordi Évole: *Igual, igual, igual...*
 Artur Mas: No, hombre, no.
 Jordi Évole: *...cuando... cuando emitamos esta entrevista, que será a partir de las diez de la noche...*
 Artur Mas: No creo, no creo que lleguen a este extremo. [...]
 Jordi Évole: Una de las cosas que se le critican es que se ha volcado mucho en la consulta. *Igual se ha olvidado mucho de otros problemas de los ciudadanos catalanes que...*
 Artur Mas: Me diga cuáles, porque esto es una acusación genérica.
- (39) Jordi Évole: Usted se sigue considerando un verso libre dentro del Partido Popular.
 Esperanza Aguirre: No, yo lo que soy es liberal, yo creo que los principios y los valores son más importantes que todo lo demás, incluido el partido.
 Jordi Évole: *¿Pero usted es consciente que a veces algunas de sus opiniones igual a Mariano Rajoy no le gustan?*
 Esperanza Aguirre: Pues es posible, es posible, pero yo he procurado no expresar aquellas discrepancias que muchas veces he tenido.

Para finalizar este epígrafe, indicamos brevemente la presencia de verbos comodín en una buena parte de las entrevistas, verbos que, como ya hemos señalado al comienzo, constituyen un rasgo de lengua coloquial. La semántica específica se pierde, dando lugar en muchos casos a errores o inexactitudes. En nuestro corpus de estudio observamos que el empleo de estos verbos (sobre todo ‘hacer’, ‘poner’ y, en menor medida, ‘tener’) se da principalmente en boca de los políticos entrevistados. El periodista utiliza “hacer el referéndum”, “tener un discurso” y “poner un impuesto”, por ejemplo, mientras que en el caso de los políticos la lista se amplía considerablemente (“hacer el referéndum”, “hacer gobierno”, “hacer un juicio”, “hacer un concurso”, “hacer una entrevista”, “hacer reformas”, “hacer una consulta”...).

7. Conclusiones

El tipo de entrevista propuesto por Jordi Évole en su programa *Salvados* ha revitalizado este género periodístico y ha logrado una amplia aceptación entre la audiencia. Uno de los aspectos que mayor atractivo presenta para los espectadores es la desenvoltura de la propia interacción conversacional entre periodista y entrevistado, en parte por el empleo de una lengua muy cercana al registro coloquial, principalmente dirigida a lograr la máxima espontaneidad durante el curso de la conversación. A pesar de tratarse de un tipo de intercambio comunicativo donde existe un reparto asimétrico del poder interaccional, se intenta eliminar esta jerarquía connatural a tal género periodístico de modo que el personaje invitado conciba la entrevista como una conversación hasta cierto punto inocente y natural, aun cuando sabe que las preguntas de Évole estarán en buena medida cargadas de ironía y de segundas intenciones. Para lograr esa espontaneidad, uno de los recursos puestos en práctica por el entrevistador es la introducción de modo evidente de un registro lingüístico coloquial, patente a través del uso de un léxico determinado, que a veces entra de lleno, además, en el terreno de lo malsonante ('joder', 'hostia'), como hemos podido constatar. El empleo en nuestro estudio de la metodología de la lingüística de corpus nos permite afirmar que algunos vocablos pertenecientes al registro coloquial adquieren incluso la consideración de palabras clave en el discurso del periodista (es el caso de 'tío' y de 'hostia'), mientras que otras palabras y expresiones propias de este mismo nivel de lengua aparecen a cada paso también en el discurso de los entrevistados, lo cual supone por parte de estos últimos una clara e implícita aceptación de tal modelo de entrevista, donde prima la tendencia a la espontaneidad y a lo coloquial en aras de un mayor acercamiento a la lengua cotidianamente utilizada por la mayor parte de la audiencia del programa.

Bionote: Andrés Ortega Garrido es *ricercatore* en la Università degli Studi di Bergamo, donde enseña lengua española en el grado de Lenguas y literaturas extranjeras modernas. Doctor en Filología y licenciado en Filología Hispánica y Filología Clásica por la Universidad Complutense de Madrid, ha enseñado en la Universidad Internacional de La Rioja, en la Università degli Studi di Milano y en el Instituto Cervantes de Milán. Autor de diferentes trabajos sobre las literaturas hispánicas, como *Vanguardia y mundo clásico grecolatino en España* (Iberoamericana, 2012) y *Géneros y tópicos de la literatura grecolatina en la poesía española de vanguardia* (Verbum, 2021), y de la edición y traducción de Petrarca, *Cartas a los más ilustres varones de la Antigüedad* (Renacimiento, 2014), se ocupa asimismo de temas inherentes a la lingüística, especialmente al análisis del discurso, como en la monografía *Estrategias discursivas en la entrevista periodística española. Salvados de Jordi Évole desde la óptica de la lingüística de corpus* (Verbum, 2022).

Author's address: andres.ortega@unibg.it



Bibliografía

- Arfuch L. 1995, *La entrevista, una invención dialógica*, Paidós, Barcelona.
- Baker P. 2006, *Using Corpora in Discourse Analysis*, Continuum, London.
- Briz Gómez A. 1996, *El español coloquial: situación y uso*, Arco Libros, Madrid.
- Briz Gómez A. 2001, *El español coloquial en la conversación. Esbozo de pragmatología*, Ariel, Barcelona.
- Cantavella Blasco J. 2014, *Análisis pragmático de una entrevista regia*, en “Estudios sobre el mensaje periodístico” 20 [1], pp. 313-329.
- Cuenca M^a J. 2013, *Uso del vocativo en la entrevista política: género discursivo y (des)cortesía*, en “Discurso & Sociedad” 7 [3], pp. 522-552.
- Escandell M^a.V. 2006, *Introducción a la pragmática*, Ariel, Barcelona.
- Fernández Lagunilla M. 2009, *La lengua en la comunicación política I. El discurso del poder*, Arco Libros, Madrid.
- Fernández Lagunilla M. 2014, *La lengua en la comunicación política II. La palabra del poder*, Arco Libros, Madrid.
- Fuentes Rodríguez C. 2013 (ed.), *Imagen social y medios de comunicación*, Arco Libros, Madrid.
- Gabrielatos C. 2018, *Keyness analysis: nature, metrics and techniques*, en Taylor C. y Marchi A. (eds.), *Corpus Approaches to Discourse: A critical review*, Routledge, London and New York, pp. 225-258.
- Garfinkel, H. y Sacks, H. 1970, *On Formal Structures of Practical Action*, en McKinney J. C. y Tiryakian E. A. (eds.), *Theoretical Sociology*, New York, Appleton Century Crofts, pp. 338-366.
- Linell P. y Luckmann T. 1991, *Asymmetries in dialogue: Some conceptual preliminaries*, en Markova I. y Foppa K. (eds.), *The Dynamics of Dialogue*, Harvester Wheatsheaf, Hemel Hempstead, pp. 1-20.
- Martínez Vallvey F. 1995, *La entrevista periodística desde el punto de vista conversacional*, Publicaciones Universidad Pontificia de Salamanca, Salamanca.
- Matte Bon F. 2011, *Gramática comunicativa del español. De la idea a la lengua. Tomo II*, Edelsa, Madrid.
- Orletti F. 2000, *La conversazione diseguale: potere e interazione*, Carocci, Roma.
- Ortega Garrido A. 2021, *Entrevista periodística, subjetividad y público implícito: estrategias discursivas en el programa Salvados de Jordi Évole*, en Meso Ayerdi K., Peña Fernández S. y Larrondo Ureta A. (eds.), *Desinformación y credibilidad en el ecosistema digital. Actas del XII Congreso Internacional de Ciberperiodismo*, Universidad del País Vasco, Servicio Editorial, Bilbao, pp. 346-363.
- Ortega Garrido A. 2022a, *Dominancia discursiva en entrevistas televisivas españolas: el caso de Esperanza Aguirre en el programa Salvados*, en Mariottini L. y Palmerini M. (eds.), *Estudios de lingüística hispánica. Teorías, datos, contextos y aplicaciones*, Dykinson, Madrid, pp. 709-729.
- Ortega Garrido A. 2022b, *Estrategias discursivas en la entrevista periodística española. Salvados de Jordi Évole desde la óptica de la lingüística de corpus*, Verbum, Madrid.
- Phillips M.A. 1989, *Lexical Structure of Text*, English Language Research, University of Birmingham, Birmingham.
- Romero Gualda M^a.V. 2008, *El español en los medios de comunicación*, Arco Libros, Madrid.
- Sacks H., Schegloff E. y Jefferson G. 1974, *A simplest systematics for the organization of turn taking for conversation*, en “Language” 50 [4], pp. 696-735.
- Scott M. 1997, *PC Analysis of Key Words And Key Key Words*, en “System” 25 [2], pp. 233-245.
- Tusón Valls A. 1997, *Análisis de la conversación*, Ariel, Barcelona.

EINE DATENGESTEUELTE UNTERSUCHUNG DEUTSCHER FUNKTIONSVERBGEFÜGE Artikelvarianz in *stehen*-Funktionsverbgefügen

JENS FLEISCHHAUER

INSTITUT OF LINGUISTICS, HEINRICH-HEINE-UNIVERSITY DÜSSELDORF

Abstract – In the paper, I present a corpus-based study on putative restrictions on articles within German light verb constructions. It is commonly proposed that the nominal element within a light verb construction is either used without any article or that the definite article is cliticized onto a preceding preposition. Based on data collected from the German reference corpus, I show that this restriction does not hold.

The analysis is based on a sample of 50 light verb constructions headed by the light verb *stehen*. I argue that light verb constructions are basically unrestricted concerning the PP-internal article, but light verb constructions show different preferences concerning article use. Thus, general claims concerning a restriction on article variance within German light verb constructions cannot be supported by actual language data. On the contrary, the result of the study reveals that restrictions on the article do neither present a criterion for identifying light verb constructions nor for distinguishing them from other types of predicative constructions.

Keywords: light verb constructions; articles; German; corpus linguistics.

1. Einleitung

Funktionsverbgefüge (FVGs) stellen seit den 1960er Jahren ein kontinuierliches Forschungsthema der germanistischen Sprachwissenschaft dar. Obwohl in verschiedenen Arbeiten (u.a. Engelen 1968; Helbig 1984; von Polenz 1987) Kriterien zur Identifikation von FVGs sowie ihrer Abgrenzung von anderen prädikativen Konstruktionen vorgeschlagen wurden, gibt es bislang keinen Konsens bezüglich ihrer Definition (siehe van Pottelberge 2001 für eine kritische Diskussion des Begriffs der Funktionsverbgefüge). Trotz des fehlenden Konsenses werden in der Forschungsliteratur einige sprachliche Merkmale als weitgehend charakteristisch für FVGs angenommen. Dazu gehören, unter anderem, (i.) dass, das in einem FVG vorkommende Nomen ein Verbalabstraktum ist, und (ii.) dass, das nominale Element nicht referenziell ist.

FVGs setzen sich zusammen aus einem Funktionsverb und einer phrasalen Komponente, die entweder eine Nominal- (1a) oder eine Präpositionalphrase (1b) sein kann. Mit der Bezeichnung ‚nominales Element‘

ist das in der phrasalen Komponente des FVGs vorkommende Nomen – *Frage* in (1a) und *Stehen* in (1b) – gemeint.

- (1a) eine Frage stellen
- (1b) zum Stehen kommen

Verknüpft mit der Idee, dass das nominale Element nicht referentiell ist, ist die Behauptung, dass ihm bestimmte Eigenschaften referenziell gebrauchter Nomen fehlen. So soll es u.a. nicht anaphorisierbar sein (z.B. Zifonun *et al.* 1997, S. 54, 1113), eine eingeschränkte Artikelvarianz aufweisen (z.B. Zifonun *et al.* 1997, S. 53, 1113; Eisenberg 2013, S. 308; Smirnova, Stöber 2022, S. 124) und nur eingeschränkt modifiziert werden können (z.B. Zifonun *et al.* 1997, S. 1113; Eisenberg 2006, S. 315; Smirnova, Stöber 2022, S. 124). Auch wenn diese Ansicht in der (germanistischen) Sprachwissenschaft verbreitet ist, gibt es bislang keine Studien, die zeigen, dass sie auch tatsächlich auf ein größeres Datensample zutreffen. Im Gegenteil, in der Regel gründen solche Behauptungen auf einer kleinen Anzahl introspektiv gewonnener Sprachdaten, die als Ausgangspunkt für Verallgemeinerungen über eine wie auch immer abgegrenzte Klasse der Funktionsverbgefüge dienen.

In der vorliegenden Arbeit soll die Behauptung, dass FVGs eine eingeschränkte Artikelvarianz aufweisen, anhand korpuslinguistischer Daten überprüft werden. In der Untersuchung geht es konkret um die Frage, ob die in den FVGs vorkommenden Nomen nur eine eingeschränkte Artikelvarianz – entweder einen klitisierten definiten Artikel oder keinen Artikel – aufweisen. Basierend auf Sprachdaten, die ich aus dem Deutschen Referenzkorpus (DeReKo, Leibniz-Institut 2021) extrahiert habe, zeige ich, dass die im Aufsatz untersuchten FVGs (i.) prinzipiell eine vollständige Artikelvarianz aufweisen und (ii.) verschiedene Faktoren (u.a. der semantische Typ des nominalen Elements) Einfluss auf die Artikelvarianz aufweisen. Aus Platzgründen kann ich mich in der vorliegenden Arbeit nicht mit der Funktion der in FVGs vorkommenden Artikel auseinandersetzen.

In Kapitel 2 gehe ich ausführlicher auf den Begriff der Funktionsverbgefüge ein. Mit der in der Forschungsliteratur behaupteten Einschränkung der FVGs bzgl. ihrer Artikelvarianz setze ich mich in Kapitel 3 auseinander. In Kapitel 4 führe ich das Untersuchungssample ein, das der korpuslinguistischen Untersuchung der Artikelvarianz, die in Kapitel 5 beschrieben wird, zugrunde liegt. Die vorliegende Arbeit setzt sich von anderen Untersuchungen zu FVGs davon ab, dass das Untersuchungssample klar definiert und empirisch durch eine Korpusstudie gewonnen wurde. Das Vorgehen zur korpuslinguistischen Untersuchung der Artikelvarianz sowie die Ergebnisse werden in Kapitel 5 berichtet. In Kapitel 6 setze ich die Ergebnisse in den Kontext der relevanten Forschungsliteratur, bevor ich mit einem Fazit in Kapitel 7 ende.

2. Funktionsverbgefüge und ihre Charakteristika

Der Begriff ‚Funktionsverb‘ wurde von Polenz (1963) eingeführt, die Bezeichnung ‚Funktionsverbgefüge‘ geht auf Engelen (1968) zurück.¹ Beide Arbeiten spielten eine wichtige Rolle in der Etablierung dieses Typs komplexer Prädikate als Untersuchungsgegenstand der germanistischen Sprachwissenschaft. Insbesondere Engelen leistete einen wichtigen Beitrag in der Herausarbeitung von Kriterien, die FVGs von anderen Prädikationstypen – z.B. regulären Prädikat-Argument-Konstruktionen (*vor dem Haus stehen*) und idiomatischen Ausdrücken (*auf dem Schlauch stehen*) – unterscheiden. Charakteristisch ist, so Engelen (1968, S. 289), dass sich die prädikative Bedeutung des FVGs aus seinem nominalen Element ergibt (auch u.a. bei Zifonun *et al.* 1997, S. 1068; Smirnova, Stöber 2022, S. 123). Dies ist evident, wenn man die wörtliche Bedeutung von *kommen* (2a) mit seiner Verwendung in dem FVG *ins Rollen kommen* (b) vergleicht. In der wörtlichen Verwendung wird eine bzgl. ihrer Art und Weise unspezifische Bewegung zu einem (deiktischen) Ziel bezeichnet. Die Bezeichnungen verschiedener Medien stehen in (2a) metonymisch für Vertreter dieser Medien, die sich zu einem Ort, der durch die NP *die Steglitzer Schule* bezeichnet ist, bewegt haben. *Grübeln* in (2b) bezeichnet keinen Ort, somit ist eine directionale Bewegungslesart für *ins Grübeln kommen* ausgeschlossen. Vielmehr kann der komplexe Ausdruck als ‚zu grübeln beginnen‘ paraphrasiert werden. Durch das Funktionsverb wird auf das Einsetzen dieser Handlung fokussiert, es trägt somit eine inchoative Bedeutung bei (z.B. von Polenz 1987, S. 170; Zifonun *et al.* 1997, S. 704).² Die Art der Handlung – *Grübeln* – wird durch das nominale Element beigetragen.

- (2a) ZDF, Sat1, RTL, Bild, die Berliner Zeitungen, Super Illu, der Hessische Rundfunk, Deutschlandfunk und sogar eine japanische Zeitung kamen in die Steglitzer Schule.
(T01/JUL.32864 die tageszeitung, 11.07.2001, S. 19, Ressort: Berlin; Spott der anderen schweißt zusammen)
- (2b) Wer bei dieser Frage ins Grübeln kommt, war bei der Firma „PR marine“ in Saal an der Donau an der falschen Adresse.
(U00/MAR.06169 Süddeutsche Zeitung, 31.03.2000, S. 48; Luxus-Werft ist pleite)

Tauscht man in dem FVG in (2b) das Nomen *Grübeln* gegen z.B. das Nomen *Stottern* aus, ändert sich die prädikative Bedeutung. *Ins Stottern kommen*

¹ Für einen historischen Überblick über die Forschung zu Funktionsverbgefügen ist auf Heine (2020) verwiesen.

² Smirnova & Stöber (2022) argumentieren, dass *kommen*-FVGs nicht immer eine inchoative Interpretation aufweisen. Da die *kommen*-FVGs nicht im Fokus der vorliegenden Untersuchung stehen, soll an dieser Stelle nicht weiter auf diesen Aspekt eingegangen werden.

bedeutet ‚zu stottern beginnen‘. Zwar liegt wieder eine inchoative Lesart vor, die Art der Handlung variiert jedoch bedingt durch die Ersetzung des FVG-internen Nomens. Anders sieht es für die Vollverbverwendung von *kommen* aus. Eine Ersetzung der NP *die Steglitzer Schule* durch *die Heinrich-Heine-Universität* ändert nicht die prädikative Bedeutung, lediglich die Spezifikation des Ziels der direktionalen Bewegung ist eine andere.

Aus dem Umstand, dass das nominale und nicht das verbale Element die prädikative Bedeutung des FVGs determiniert, leitet Engelen weitere Eigenschaften der FVGs ab: (i) das Nomen kann nicht erfragt werden, (ii) das Nomen kann nicht pronominalisiert werden, (iii) das Nomen kann nicht durch einen Infinitiv ersetzt werden, (iv) das Nomen ist ein Verbalabstraktum und (v) das Nomen ist nicht attribulierbar.

Engelen begründet nicht, warum das nominale Element ein Verbalabstraktum, also ein denominales Verb, sein soll. Von Polenz greift dies auf und schreibt:

Das Substantiv in FVG [...] ist satzsemantisch ein Prädikatsausdruck, d.h. Bezeichnung eines Prädikats, und zwar einer HANDLUNG, TÄTIGKEIT, eines VORGANGS, EREIGNISSES oder ZUSTANDS [...], die den prädikativen Kern des FVG darstellt. Mit dem [Substantiv] wird also nicht auf etwas referiert/bezogen [...] (von Polenz 1987, S. 175)

Von Polenz kreiert eine Opposition zwischen Prädikation und Referenz, wobei er die Bezeichnung eines Ereignisses oder Zustands als Funktion eines (satzsemantischen) Prädikatsausdrucks ansieht.

Eine strikte Opposition zwischen Referenz und Prädikation ist in dieser Form aber nicht gegeben. Aus semantischer Perspektive stellen Nomen genauso Prädikate dar wie Verben und Adjektive auch. Ein Nomen wie *Hund* prädiziert, dass der Referent unter die Prädikation ‚Hund‘ fällt und somit bestimmte Charakteristika aufweist (z.B. zur Gattung ‚Hund‘ gehörend, fleischfressend, bellend). Erst durch eine referenzielle Verankerung – beispielsweise durch die Kombination mit dem in-/definiten Artikel – wird ein Nomen referenziell gebraucht. Der referenzielle Gebrauch kann beispielsweise durch Anaphern getestet werden. Durch *den Hunden* wird in (3) ein Referent eingeführt, der durch das Demonstrativum *diese* wieder aufgegriffen werden kann.

(3) Der Junge spielte mit den Hunden_i. Diese_i haben braunes Fell.

Spätestens seit den Arbeiten Donald Davidsons (1980, 1985) ist die Ansicht, dass auch (die meisten) Verben referieren, weit verbreitet. Im Rahmen der Ereignissemantik (z.B. Davidson 1980, 1985; Parsons 1990) weisen Verben ein Ereignisargument auf. Evidenz für die Annahme eines solchen Arguments ist unter anderem, dass das vom Verb eingeführte Ereignis anaphorisch

aufgenommen werden kann (u.a. Maienborn 2011). In (4) verweist das anaphorisch gebrauchte Demonstrativum auf das von dem Satz bezeichnete Ereignis: was gestern im Garten geschah war ein Spielereignis in dem ein Junge und ein Hund involviert waren.

(4) Der Junge spielte mit den Hunden. Dies geschah gestern im Garten.

Prädikation und Referenz schließen sich also gegenseitig nicht aus. In der relevanten Forschungsliteratur zu Funktionsverbgefügen ist dennoch die Referenzlosigkeit des FVG-internen Nomens – explizit oder implizit – eine wiederkehrende zentrale Annahme (z.B. auch bei Eisenberg 2013, S. 308f.). Aus dieser postulierten Referenzlosigkeit ergeben sich konkrete Folgerungen bezüglich der semantischen und grammatischen Eigenschaften des Nomens. Nach von Polenz weisen die Nomen folgende Eigenschaften auf, die sich aus ihrer nicht-Referenzialität ableiten sollen: (i) nicht erfragbar, (ii) nicht pronominalisierbar, (iii) meist im Singular (Verlust der Numerusopposition), (iv) kaum Artikelvariation. Partiiell sind diese Eigenschaften identisch mit denen, die Engelen als charakteristisch für FVGs ansieht.

Inwiefern diese (und andere) Eigenschaften tatsächlich auf FVGs zutreffen, ist bislang nicht auf Grundlage einer größeren Datenbasis untersucht worden. In insbesondere der älteren Forschungsliteratur sind solche Charakteristika auf Basis einzelner ausgewählter Beispiele postuliert worden, die dann als Eigenschaften aller FVGs verallgemeinert wurden.³

Die vorliegende Arbeit hat zum Ziel, eine der postulierten Eigenschaften – nämlich die Beschränkung der Artikelvarianz – auf Grundlage eines definierten Samples korpuslinguistisch zu überprüfen. Bevor ich das Untersuchungssample in Kapitel 4 vorstelle, soll zunächst konkreter auf die angenommenen Beschränkungen der Artikelvarianz in FVGs eingegangen werden.

3. Artikelvarianz in FVGs

Das Deutsche verfügt über eine Opposition zwischen definitem und indefinitem Artikel, die in freier (5a/b) und klitisierter Form (5c/d) vorkommen können. Individualnomen, wie etwa *Herzinfarkt*, weisen in der Regel freie Artikelvarianz auf (vgl. 5). Das bedeutet, dass Individualnomen mit freien und klitisierten definiten und indefiniten Artikeln vorkommen können. In

³ Einige Autoren haben nuancierte Aussagen vorgenommen und unterscheiden beispielsweise zwischen echten und unechten, bzw. lexikalisierten und nicht-lexikalisierten FVGs (z.B. Helbig 1984), wobei die postulierten Eigenschaften nur für eine Teilklasse (z.B. die echten/lexikalisierten FVGs) zutreffen muss, während die andere Teilklasse die postulierten Eigenschaften nicht notwendigerweise aufweist.

bestimmten Kontexten ist auch eine Verwendung ohne Artikel möglich, sogenannte ‚nackte Plurale‘ wie in (5e).⁴

- (5a) So gilt der Herzinfarkt fälschlicherweise als Männerkrankheit, obwohl Frauen ihn genauso häufig erleiden – allerdings eher in einem höheren Lebensalter. (STG21/SEP.00041 Stern Gesund leben, 15.09.2021, S. 62; DIE ERFORSCHUNG DER FRAU)
- (5b) Offenbar kostete ihm ein Herzinfarkt das Leben. (NON07/NOV.06237 Niederösterreichische Nachrichten, 13.11.2007, S. 5; Tod im Einkaufszentrum)
- (5c) „Wir waren alle kurz vorm Herzinfarkt“, kommentierte Trainer Jürgen Klopp die historische Aufholjagd. (BRZ13/APR.03841 Braunschweiger Zeitung, 11.04.2013, Ressort: 1WN-Spo; Zorc: Die aufregendsten 120 Sekunden meines Lebens)
- (5d) Mein Herz schlägt dermaßen schnell, dass ich kurz vor'nem Herzinfarkt steh.⁵
- (5e) „Herzinfarkte sind schon lange keine Managerkrankheit mehr“, weiß Carmen Bresny. (NKU12/OKT.06606 Nordkurier, 19.10.2012; Stressige Jobs führen öfter zu Herzkrankheiten)

Artikelvarianz in FVGs wird in der Forschungsliteratur in der Regel *en passant* abgehandelt, meist auf Grundlage weniger introspektiv gewonnener Beispiele. Repräsentativ ist die Behandlung der Artikelvarianz bei Leiss (2000). Sie führt die Beispiele in (6) an, von denen das zweite ihrer Einschätzung nach nicht akzeptabel sei. Daraus schlussfolgert sie, dass die Artikelvarianz neutralisiert ist und das FVG-interne Nomen nur mit klitisiertem Artikel vorkommt.⁶

- (6a) Er bringt das Stück zur Aufführung.
- (6b) *Er bringt das Stück zu einer Aufführung.
(Leiss 2000, S. 208)

Dies scheint die Mehrheitsposition bezüglich des Vorkommens von Artikeln in FVGs zu sein und findet sich ähnlich formuliert z.B. bei Helbig (1984), von Polenz (1987), Zifonun *et al.* (1997, S. 53), Fuhrhop (2007, S. 275), Eisenberg (2013, S. 308) und Smirnova & Stöber (2022, S. 124). Von Polenz (1987, S.

⁴ Massennomen wie *Sand* und *Gold* weisen Einschränkungen in der Artikelvarianz auf und sind mit dem Indefinitartikel nicht kompatibel. Eine solche Einschränkung ergibt sich aufgrund ihrer semantischen Eigenschaften, vgl. Krifka (1991).

⁵ <https://context.reverso.net/%C3%BCbersetzung/deutsch-englisch/kurz+vor+%27nem>; Aufruf: 27.02.2023.

⁶ Tatsächlich ist das Vorkommen des freien indefiniten Artikels innerhalb des FVGs *zu Aufführung bringen* im DeReKo belegt, wie folgendes Beispiel zeigt:

- (i) Mit seinem Star-Quintett (Maria Nicklisch, Friedrich Domin, Wilfried Seyferth, Peter Lühr und Charles Regnier) gelang es ihm, Dürrenmatts Zwitter ("zwischen Pamphlet und Kunstwerk") zu einer Aufführung zu bringen, die Kritiker-Attribute wie "ungewöhnlich" und "außergewöhnlich" rechtfertigte. (S52/APR.00034 Der Spiegel, 02.04.1952, S. 31; Ideen fressen Menschen)

Dies untermauert noch einmal, wie kritisch Verallgemeinerungen auf der Grundlage einzelner introspektiv gewonnener Beispiele sind.

171) schreibt, dass das FVG-interne Nomen meist ohne Artikel oder mit einem an die Präposition klitisierten Artikel vorkommt (ebenso Herrlitz 1973, S. 21). Eine vorsichtiger Formulierung bezüglich der eingeschränkten Artikelvarianz findet sich bei Helbig (1984, S. 170), der schreibt, dass es „auffällige Restriktionen im Artikelgebrauch“ gibt. Nach ihm ist der Artikel zumeist festgelegt und entweder ein Nullartikel – also kein overt Artikel – oder ein definitiver Artikel, der an die Präposition klitisiert ist.

Im Unterschied zu von Polenz konstatiert Helbig jedoch, dass es Ausnahmen zu dieser strikten Artikelfixierung gibt. Er führt die Beispiele in (7) an, die freie Artikelvariation illustrieren. Diese Beispiele enthalten jedoch nur Konstruktionen des Typs ‚Funktionsverb + NP‘, die von Polenz (1984, S. 169ff.) als ‚Nominalisierungsverbgefüge‘ aus dem Phänomenbereich der Funktionsverbgefüge ausgegrenzt. Unklar ist damit, welche Position Helbig bezüglich der Artikelvariation von FVGs des Typs ‚Funktionsverb + PP‘ einnimmt.

- (7a) Er nimmt Kontakt/ den Kontakt/ einen Kontakt mit seinen Genossen auf.
 - (7b) Wir haben ihn erst auf den/ einen Gedanken gebracht.
 - (7c) Wir geben ihm Antwort/ die Antwort/ eine Antwort.
- (Helbig 1984, S. 173)

Helbig verweist in seiner Diskussion auf den ‚Nullartikel‘, also eine phonologisch nicht overte Artikelrealisierung. Ein Nullartikel steht nicht nur in Opposition zu overt Artikelrealisierungen, sondern auch zur Weglassung des Artikels. In der vorliegenden Arbeit setze ich keinen Nullartikel an, sondern werte das Fehlen eines overt Artikels als keine Artikelrealisierung.

Abgeleitet aus den oben skizzierten Forschungspositionen kann folgende Hypothese festgehalten werden:

- (8) Das in Funktionsverbgefügen vorkommende PP-interne Nomen tritt entweder ohne Artikel oder mit einem an die Präposition klitisierten definiten Artikel auf.

Die Hypothese ist explizit auf Funktionsverbgefüge des Typs ‚Funktionsverb + PP‘ eingeschränkt, Nominalisierungsverbgefüge sind in dieser Hypothese sowie auch der weiteren Untersuchung nicht eingeschlossen. Die Hypothese in (8) schließt insbesondere aus, dass in Funktionsverbgefügen der definite und indefinite Artikel frei vorkommt. Sollte sich diese Vermutung erhärten, dann würden sich die FVGs bezüglich Artikelvarianz unter anderem von den Nominalisierungsverbgefügen, wie in (7), deutlich unterscheiden. Diese Hypothese ist insoweit relevant, als verschiedene Autoren, u.a. Eisenberg (2006, S. 315) und Fuhrhop (2007) die in (8) formulierte Restriktion bzgl. der

Artikelvarianz als charakteristische Eigenschaft deutscher FVGs ansehen.⁷ Für die weitere Untersuchung ergibt sich daraus die Forschungsfrage, ob die in (8) formulierte Restriktion bzgl. der Artikelvarianz anhand korpuslinguistischer Daten nachgewiesen werden kann. Die korpuslinguistische Untersuchung der Artikelvarianz ist Gegenstand von Kapitel 5, zuerst führe ich in Kapitel 4 das Untersuchungssample ein.

4. Untersuchungssample: *stehen*-FVGs

In Kapitel 2 wurde bereits erwähnt, dass es bezüglich einer Definition des Begriffs der Funktionsverbgefüge bislang keinen Konsens in der Forschungsliteratur gibt. Somit besteht auch kein Konsens, welche Ausdrücke Funktionsverbgefüge darstellen und welche nicht. Für die vorliegende Arbeit greife ich auf ein Untersuchungssample von *stehen*-FVGs zurück, das in verschiedenen Arbeiten (Fleischhauer 2021a, 2021b, 2022, 2023a, 2023b) auf Grundlage korpuslinguistischer Untersuchungen erstellt wurde. Das diesen Untersuchungen zugrundeliegende korpuslinguistische Vorgehen wird in Kapitel 4.2 dargestellt. Zunächst führe ich in Kapitel 4.1 einige grundlegende Beobachtungen zu *stehen*-FVGs an.

4.1. *stehen*-FVGs und ihre Familien

Laut Kamber (2006) ist *stehen* eines der frequentesten Funktionsverben im Deutschen (auch Eisenberg 2006, S. 307 verweist auf seine besondere Häufigkeit). Als lexikalisches Vollverb bezeichnet *stehen* einen Zustand, in dem der Subjektreferent in einer aufrechten Körperhaltung an einer bestimmten räumlichen Position statisch lokalisiert ist (z.B. Gamerschlag *et al.* 2013). Die Position des Subjektreferenten (*der Hund*) wird in (9) durch das PP-Komplement des Verbs (*vor dem Haus*) als sich in einer Vor-Region des Hauses befindlich spezifiziert (zum Begriff der ‚Vor-Region‘ ist auf Wunderlich, Herweg 1991 verwiesen).

(9) Der Hund steht vor dem Haus.

Im Vergleich zur Vollverbverwendung ist das Funktionsverb desemantisiert. Der Subjektreferent in (10) muss sich nicht in aufrechter Position statisch an einer Stelle befinden, vielmehr kann er sich sogar bewegen, während er unter Beobachtung steht.

⁷ Harm (2021; Kap. 6.3) zeigt, dass die in (9) formulierte Hypothese für eine Reihe deutscher FVGs nicht zutrifft. Die vorliegende Arbeit unterscheidet sich von der Harms durch die Wahl der untersuchten FVGs sowie durch ihre quantitative Ausrichtung.

- (10) Der Verdächtige steht (während seiner Reise nach Frankreich) unter Beobachtung.

Auch wenn das Funktionsverb desemantisiert ist, ist es dennoch nicht bedeutungsleer. Dies zeigt sich etwa in Kommutationsreihen (Helbig 1984; von Polenz 1987; Zifonun *et al.* 1997, S. 704), wie in (11). Die drei Funktionsverbgefüge variieren lediglich bezüglich des Funktionsverbs, die PP in den Beispielen ist jeweils identisch. *Unter Beobachtung stellen* (11a) hat eine kausative Bedeutung, *unter Beobachtung geraten* (11b) eine inchoative und *unter Beobachtung stehen* (11c) eine stativische. Da nur das Funktionsverb variiert, ist es plausibel anzunehmen, dass der Bedeutungsunterschied auf das verbale Element zurückzuführen ist.

- (11a) Im Laufe des 31. Mai wurde bekannt, daß belgische Behörden auch Untersuchungen an Schweinefleisch veranlaßt haben. Am 3. Juni haben belgische Behörden vorsorglich die Schlachtung von Schweinen ausgesetzt und neben 500 Schweinemastbetrieben auch 70 Rinderbetriebe unter Beobachtung gestellt [...]
(PBY/W14.00019 Protokoll der Sitzung des Parlaments Bayerischer Landtag am 09.06.1999. 19. Sitzung der 14. Wahlperiode 1998-2003. Plenarprotokoll, München, 1999)
- (11b) Hunderte Millionen von Menschen seien im Zusammenhang mit der Nutzung von Internet, Telefon, E-Mail oder sozialen Netzwerken unter Beobachtung geraten.
(NZZ14/JAN.00246 Neue Zürcher Zeitung, 04.01.2014, S. 23; Protest gegen Überwachung)
- (11c) Der FCB stand sowieso schon unter Beobachtung und ist nun endgültig gebrandmarkt.
(E14/MAR.02376 Tages-Anzeiger, 27.03.2014, S. 48; Am Ende der Geduld)

Anders als etwa Zifonun *et al.* (1997, S. 1113) nehme ich nicht an, dass die Präposition „vergleichsweise desemantisiert“ ist. Vielmehr trägt die Präposition systematisch zur Bedeutung bei, da die Interpretation der *stehen*-FVGs unter anderem von der Präposition abhängt.⁸ Das FVG *unter Beobachtung stehen* (11c) hat eine passivische Interpretation und kann als ‚beobachtet werden‘ paraphrasiert werden. *Vor einer Beobachtung stehen*, in (12), hat dagegen eine prospektive Interpretation (Bogaards, Fleischhauer 2023; Fleischhauer 2023a; Fleischhauer, Gamerschlag 2019; Fleischhauer *et al.* 2019) und ist paraphrasierbar als ‚davor sein beobachtet zu werden‘.

⁸ Der figurative semantische Beitrag zu den hier behandelten FVGs wird ausführlicher in Fleischhauer (2021b, 2022) für *unter* und Fleischhauer & Gamerschlag (2019) sowie Fleischhauer *et al.* (2019) für *vor* diskutiert,

- (12) Denn auch die AfD steht möglicherweise vor einer Beobachtung durch den Inlandsgeheimdienst.
(L18/NOV.03161 Berliner Morgenpost, 26.11.2018, S. 5; AfD-Nachwuchs vor der Spaltung?)

Die beiden FVGs in (11c) und (12) weisen neben unterschiedlichen Paraphrasen, bedingt durch die verschiedenen Präpositionen, noch weitere Unterschiede auf. So können die passivischen FVGs unter anderem mittels *stellen* kausativiert werden (13a), die prospektiven FVGs dagegen nicht (13b).

- (13a) Eine neue Variante des Coronavirus zieht immer mehr die Aufmerksamkeit auf sich, nachdem die Weltgesundheitsorganisation (WHO) sie unter Beobachtung gestellt hat.
(L21/AUG.00465 Berliner Morgenpost, 05.08.2021, S. 3; Wie gefährlich ist Lambda?)
- (13b) #Das Innenministerium stellt die AfD vor eine Beobachtung durch den Inlandsgeheimdienst.

Ein erster Vergleich bezüglich der Artikelsetzung zeigt, dass das FVG in (11c) ohne Artikel verwendet wird, das in (12) dagegen einen definiten Artikel aufweist. Aus dieser Beobachtung lässt sich die Frage ableiten, ob verschiedene Typen von Funktionsverbgefügen unterschiedliche Restriktionen hinsichtlich der Artikelvarianz aufweisen?

FVGs variieren in ihrer Interpretation jedoch nicht nur in Abhängigkeit von der Präposition, auch das PP-intern vorkommende Nomen wirkt sich auf die Interpretation aus. Während *unter Beobachtung stehen* durch eine Vorgangspassivkonstruktion – ‚beobachtet werden‘ – paraphrasiert wird, ist ‚gestresst sein‘ – also eine Zustandspassivkonstruktion – die passende Paraphrase für *unter Stress stehen* (14).

- (14) Kinder, die ständig unter Stress stehen, können nämlich nicht lernen.
(FOC02/FEB.00475 FOCUS, 25.02.2002, S. 048-050; „Von Kindern keine Ahnung“)

Funktionsverbgefüge, die das gleiche Interpretationsmuster aufweisen und nur bezüglich ihres PP-internen Nomens variieren, bilden eine Familie (Fleischhauer 2022, S. 257; Fleischhauer, Turus 2022, S. 179). *Unter Beobachtung stehen* und *unter Stress stehen* gehören zwei unterschiedlichen Familien an, da sie unterschiedliche Interpretationsmuster – Vorgangs- vs. Zustandspassiv – aufweisen. Die Annahme zweier distinkter *stehen unter*-Familien wird dadurch erhärtet, dass die FVGs der entsprechenden Familien unterschiedliche grammatikalische Eigenschaften aufweisen. Während *unter Beobachtung stehen* mit *stellen* kausativiert werden kann (13a), nimmt *unter Stress stehen* das Funktionsverb *setzen* zur Kausativierung (15). Die Verwendung von *stellen* wäre in (15) inakzeptabel (vgl. Fleischhauer 2022, S.

276ff. für korpuslinguistische Daten zur Kausativierung der *stehen unter*-FVGs).

- (15) Für Athleten, die ihren Körper ständig unter Stress setzen, kann das spürbare Folgen haben.
(NZZ19/OKT.00482 Neue Zürcher Zeitung, 07.10.2019, S. 35; Julien Wanders nimmt den Hinterausgang)

Die FVGs einer Familie sind sowohl morphosyntaktisch als auch semantisch definiert. Bezüglich ihrer morphosyntaktischen Form dürfen die FVGs einer Familie nur hinsichtlich des PP-internen Nomens variieren, Funktionsverb und Präposition müssen konstant sein (Fleischhauer 2021a, S. 64; Fleischhauer 2022, S. 257; Fleischhauer, Turus 2021, S. 77; Fleischhauer, Turus 2022, S. 179). Hinsichtlich ihrer Semantik weisen alle Mitglieder einer Familie dasselbe Interpretationsmuster auf, was durch Einheitlichkeit ihrer Paraphrase aufgezeigt werden kann. Beide Eigenschaften sind für die in der vorliegenden Arbeit untersuchten FVGs in Tabelle 1 zusammengefasst.⁹ Ausdrücke des Typs ‚*stehen unter* + NP‘ stellen FVGs der Vorgangspassiv-Familie dar, wenn sie mittels einer Vorgangspassivkonstruktion (‚PARTIZIP II werden‘) paraphrasiert werden. Entsprechendes gilt für die jeweiligen FVGs der Zustandspassiv-, bzw. Prospektiv-Familie.

Familie	Morphosyntax	Paraphrase
Vorgangspassiv-Familie	<i>stehen unter</i> NP	‚PARTIZIP II werden‘
Zustandspassiv-Familie	<i>stehen unter</i> NP	‚PARTIZIP II sein‘
Prospektiv-Familie	<i>stehen vor</i> NP	‚davor sein INF‘

Tabelle 1
Zusammenfassung der untersuchten FVG-Familien.

4.2. Identifikation der *stehen*-FVGs

Da das Ziel des vorliegenden Aufsatzes eine datengesteuerte Untersuchung der Artikelvarianz in FVGs ist, muss in einem ersten Schritt das Datensample erstellt werden. Dazu eignen sich, wie im letzten Abschnitt dargelegt, FVG-Familien als Ausgangspunkt. Für die Vorgangspassiv-, Zustandspassiv- und Prospektiv-Familie liegen korpuslinguistische Untersuchungen vor (Fleischhauer 2021a, 2021b, 2022, 2023a, 2023b), deren Ziel unter anderem in einer Bestimmung der Mitglieder dieser Familien bestand. Das korpuslinguistische Vorgehen war in der Untersuchung der einzelnen Familien identisch. Für die korpuslinguistische Untersuchung wurde das Archiv W des Deutschen Referenzkorpus (DeReKo; Leibniz-Institut 2021) verwendet, das eine große Bandbreite von Texten (u.a. Zeitungen, Biografien, Wikipedia) aus

⁹ In Fleischhauer (2022, 2023a) werden weitere FVG-Familien diskutiert, die aus Platzgründen nicht in die vorliegende Untersuchung eingehen konnten.

dem Zeitraum vom 18. Jahrhundert bis zur Gegenwart enthält. Die Größe des Korpus beträgt 455,85 Millionen Wortformen (Stand November 2023) und es ist mittels der Software COSMAS-2 (Leibniz-Institut 2020) webbasiert zu durchsuchen. Im Archiv W des DeReKo wurden mittels der Suchstrings ‚&stehen /s0 vor‘ und ‚&stehen /s0 unter‘ alle Vorkommen von *stehen* – infinite, flektierte und derivierte – und der Präposition *vor*, bzw. *unter* gesucht. *Stehen* und die Präposition mussten innerhalb desselben Satzes (, \s0‘) vorkommen. Eine solche Beschränkung ist sinnvoll, da das Funktionsverb und die PP nicht adjazent stehen müssen, wie das Beispiel in (16) zeigt.

- (16) Und so steht der Kanzler immer dann, wenn er sich in den Räumlichkeiten auf der Nordseite aufhält, im kleinen Besprechungsraum, bei Kanzleramtsminister Steinmeier oder seiner Büroleiterin Sigrid Krampitz, unter Beobachtung.
(L03/NOV.00428 Berliner Morgenpost, 02.11.2003, S. 2; Er weiß genau, was da vor sich geht, drüben im Kanzleramt. Er beobachtet es schon lange durch sein großes Fernglas. Gerhard Fischer ist der Kanzler-Spion)

Die Intention hinter der Suchanfrage bestand in der Ermittlung der Varianz bzgl. des PP-internen Nomens. Auf diese Weise konnte der Umfang der FVGs der einzelnen Familien korpuslinguistisch ermittelt werden. Die Suchanfrage ‚&stehen /s0 unter‘ ergab 80255 Treffer, von denen 10% (8023 Sätze) zufällig für die Analyse ausgewählt wurden. 55 Treffer mussten aussortiert werden, da sie unvollständig oder doppelt waren. Der Suchstring ‚&stehen /s0 vor‘ erbrachte 1241876 Treffer; 10000 zufällige Sätze wurden für die Analyse extrahiert. 19 Belege mussten aussortiert werden, da sie unvollständig oder doppelt waren.

In einem ersten Annotationsschritt mussten die Fälle, in denen die Präposition Kopf einer Adjunkt-PP ist, aussortiert werden. Dazu wurde getestet, ob die PP weggelassen werden kann, ohne dass sich die Akzeptabilität des Satzes, bzw. die prädikative Bedeutung ändert. Dies ist etwa in (17) der Fall, die Weglassung der PP *unter Verkehrsvereinsmitgliedern* hat weder Auswirkungen auf die Akzeptabilität noch auf die Bedeutung des Satzprädikats. Dieses wird vielmehr durch das Gefüge *im Zentrum stehen* gebildet. Aus diesem Grund ist die PP *unter Verkehrsvereinsmitgliedern* als Adjunkt, bzw. Attribut und nicht als Bestandteil des FVGs gewertet worden. Im Falle der FVGs stellt die PP einen obligatorischen Bestandteil des komplexen Prädikats dar und ist somit kein Adjunkt, sondern ein Komplement des Funktionsverbs.

- (17) Informationen aus erster Hand sowie die Kontaktpflege unter Verkehrsvereinsmitgliedern standen im Zentrum des zweiten Tourismus-Workshops, zu dem Rapperswil Zürichsee Tourismus einlud.
(SOZ06/FEB.04063 Die Südostschweiz, 20.02.2006; Tourismus mit Know-how und Netzwerk)

	<i>stehen vor</i>	<i>stehen unter</i>
PP: Komplement	6111	5822
PP: Adjunkt	3870	2146
Gesamt	9981	7968

Tabelle 2

Resultate des ersten Annotationsschritts (vgl. Fleischhauer 2022, S. 262, Fleischhauer 2023a, S. 382).

In einem zweiten Annotationsschritt wurde die Funktionsverbverwendung von *stehen* von ihrer Vollverbverwendung unterschieden. Eines der Kriterien, das dazu angewendet wurde, war, ob das Verb durch ein anderes Verb derselben semantischen Klasse ersetzt werden kann (u.a. Heringer 1968, S. 36, 44; Helbig 1984, S. 168; von Polenz 1987, S. 172). Als Vollverb stellt *stehen* ein Verb der Körperhaltung dar (Kaufmann 1995, S. 102) und kontrastiert u.a. mit *sitzen*, *liegen*, *hocken*. In (18) kann *stehen* durch *liegen* ersetzt werden, was zeigt, dass das Verb als Vollverb verwendet ist. Die Ausrichtung des Subjektreferenten ist nicht immer im Fokus der Prädikation, sodass *stehen* auch als reines Lokationsverb verwendet wird. In dieser Verwendung ist es durch *positioniert/lokalisiert sein* substituierbar. Eine solche Substitution ist auch in (17) möglich.

- (18) Da der Korb unter dem Hausdach stand, weitete sich das Feuer rasant aus.
 (BRZ07/DEZ.19861 Braunschweiger Zeitung, 07.12.2007; Strandkorb-Brand:
 Ursache unklar)

In (19) kann *stehen* weder durch ein Verb der Körperhaltung noch durch ein statisches Lokationsverb ersetzt werden. *Stehen* wird somit als Funktionsverb gebraucht.

- (19) Am Anfang seiner Laufbahn stand er aber besonders unter dem Einfluss seines Vaters August.
 (WPD/WWW.04814 FloSch; AlexR; 5: Fritz von Wille, In: Wikipedia -
 URL:<http://de.wikipedia.org>: Wikipedia, 2005)

Der zweite Annotationsschritt wurde nur auf die Sätze, die die *vor*-, bzw. *unter*-PP als Komplement enthalten, angewendet. Die Ergebnisse sind in Tabelle 3 zusammengefasst.

	<i>stehen vor</i>	<i>stehen unter</i>
Vollverbverwendung	3222	562
Funktionsverbverwendung	2889	5260

Tabelle 3

Resultate des zweiten Annotationsschritts (vgl. Fleischhauer 2022, S. 265, Fleischhauer 2023a, S. 383).

In einem letzten Schritt konnten die Mitglieder der einzelnen Familien identifiziert werden. Das relevante Kriterium stellt, wie im letzten Abschnitt beschrieben, die Paraphrase dar. Tabelle 4 fasst die Anzahl der identifizierten Nomentypen, die in den FVGs der einzelnen Familien vorkommen, zusammen. Dies ergibt zugleich die Anzahl aller in den Korpusdaten identifizierten Mitglieder der einzelnen Familien.¹⁰ Wichtig ist, dass dies keine abschließende Erfassung aller Mitglieder der drei FVG-Familien darstellt, sondern nur die Mitglieder umfasst, die im Korpus sample repräsentiert sind.

	Vorgangspassivfamilie	Zustandspassivfamilie	Prospektivfamilie
Nomentypen	34	19	305

Tabelle 4

Anzahl der distinkten Nomentypen innerhalb der drei FVG-Familien.

5. Korpusuntersuchung der Artikelvarianz

5.1. Korpuslinguistisches Vorgehen

Mittels der im letzten Abschnitt zusammengefassten Korpusuntersuchungen wurden 358 distinkte *stehen*-Funktionsverbgefüge identifiziert. Für die korpuslinguistische Untersuchung der Artikelvarianz habe ich alle FVGs der Vorgangs- und Zustandspassivfamilien herangezogen. Für die Prospektivfamilie habe ich mich auf die zehn frequentesten FVGs, bemessen an den Korpusuntersuchung in Fleischhauer (2023a), beschränkt.

Da es um die Untersuchung der Artikelvarianz innerhalb bestimmter FVGs geht, musste der Suchstring entsprechend angepasst werden. Für die Suchanfrage wurde der String ‚&stehen /s0 P /w4 N‘ verwendet, wobei ‚P‘ für die Präpositionen *unter*, bzw. *vor* steht und ‚N‘ für das entsprechende PP-interne Nomen des jeweiligen FVGs. Für das FVG *unter Beobachtung stehen* wurde somit der Suchstring ‚&stehen /s0 unter /w4 Beobachtung‘ verwendet. Die Suchanfrage liefert alle Vorkommen von *stehen* und *unter*, die innerhalb eines Satzes vorkommen und darauf eingeschränkt sind, dass zwischen *unter* und *Beobachtung* ein maximaler Abstand von vier Wörtern besteht. Durch die Wahl eines Abstands von vier Wörtern können neben einem Artikel auch attributive Adjektive innerhalb der PP-internen Nominalphrase vorkommen.

Wie auch in den Korpusuntersuchungen, die im letzten Abschnitt dargestellt wurden, wurde wieder auf das Archiv W des DeReKo zurückgegriffen. Für jedes FVG wurde die maximale Anzahl an zufällig sortierten Treffern extrahiert. Die Daten wurden manuell annotiert; in einem

¹⁰ Eine Auflistung aller in den drei Familien vorkommenden Nomen ist online einsehbar: https://osf.io/hb5j8/?view_only=4b81b92b9d994cb6aa3088bfdae2954c.

ersten Schritt wurde getestet, ob die PP Komplement von *stehen* ist. Damit wurde der erste Annotationsschritt aus Kapitel 4.2 wiederholt. Auf diese Weise wurden für jedes FVG 100 Belege identifiziert, die im nächsten Annotationsschritt bezüglich NP-interner Adjunkte (Artikel, attributive Adjektive) annotiert wurden. Belege, in denen das FVG in einer Überschrift vorkommt, wurden aussortiert. Es ist bekannt, dass Artikel in Überschriften häufig optional sind (z.B. Sandig 1971; Reich 2017), sodass der Einbezug solcher Belege die Auswertung zu Gunsten artikelloser Vorkommen hätte beeinflussen können.

Mehrere FVGs wurden, da über das Suchkriterium keine 100 Belege extrahiert werden konnten, von der weiteren Untersuchung ausgeschlossen. In den Fällen, in denen mehr als 100 Belege gefunden wurden, erfolgte eine automatische Zufallssortierung und eine Beschränkung auf die ersten 100 Belege. Tabelle 5 gibt einen Überblick über die in der Untersuchung verwendeten FVGs, für jede Familie werden dabei nur die PP-internen Nomen angeführt. Insgesamt bestand das Untersuchungssample somit aus 25 FVGs der Vorgangspassivfamilie, 15 FVGs der Zustandspassivfamilie und 10 FVGs der Prospektivfamilie.

Familie	PP-interne Nomen
Vorgangspassivfamilie	<i>Aufsicht, Befehl, Belastung, Beobachtung, Besatzung, Beschuss, Betreuung, Bewachung, Einwirkung, Entzug, Federführung, Fittiche, Fuchtel, Führung, Herrschaft, Kommando, Kontrolle, Kritik, Leitung, Schirmherrschaft, Schutz, Strafe, Überwachung, Verdacht, Vertrag, Verwaltung</i>
Zustandspassivfamilie	<i>Alkohol, Anklage, Anspannung, Dampf, Drogen, Druck, Einfluss, Feuer, Medikamente, Schock, Spannung, Stress, Strom, Zugzwang, Zwang</i>
Prospektivfamilie	<i>Abschluss, Bankrott, Durchbruch, Kollaps, Konkurs, Pleite, Ruin, Sieg, Wechsel, Zusammenbruch</i>

Tabelle 5
 Auflistung der untersuchten FVGs.

Für die Auswertung der Untersuchung wurden alle Ausdrücke, die zwischen den Präpositionen *unter* bzw. *vor* und dem jeweiligen Kopfnomen vorkommen, berücksichtigt. Prinzipiell können zwei Arten von Ausdrücken zwischen diesen Elementen stehen: Determinierer und attributive Modifikatoren. Unter Determinierer fasse ich alle Ausdrücke, die in einer paradigmatischen

Beziehung mit Artikeln stehen und syntagmatisch mit Modifizierern kombiniert werden können (vgl. Vater 1979). Neben den definiten und indefiniten Artikeln traten in der vorliegenden Untersuchung auch verschiedene pronominale Formen, z.B. Demonstrativ- (20a), Possessiv- (20b) und Indefinitpronomen (20c), sowie Genitiv-NPs (20d) auf. Das Indefinitpronomen *kein* habe ich separat erfasst, da es laut der relevanten Literatur in FVGs nicht vorkommt (von Polenz 1987, S. 175; Helbig 1985, S. 171). Statt der PP-internen Negation mit *kein* soll in der Regel lediglich eine Prädikatsnegation mit *nicht* möglich sein. Unter die Modifikatoren fallen vor allem attributiv gebrauchte Adjektive, die mit den Determinierern kombiniert werden können. Ein weiterer Modifikatortyp, Relativsätze, kommt im Datensample nur postnominal vor und wurde nicht systematisch in der Analyse berücksichtigt (siehe Kapitel 5.4).

- (20a) Ich erkläre mir das ein Stück weit damit, dass andere Länder nicht unter dieser Belastung stehen wie die Bundesrepublik Deutschland.
(PHE/W15.00084 Protokoll der Sitzung des Parlaments Hessischer Landtag am 23.10.2001. 84. Sitzung der 15. Wahlperiode 1999-2003. Plenarprotokoll, Wiesbaden, 2001)
- (20b) So würde sie sowohl dem Unternehmen dienen als auch unter seiner Aufsicht stehen.
(U04/AUG.03525 Süddeutsche Zeitung, 21.08.2004, S. ROM7; Tatjanas Emanzipation)
- (20c) „Das zeigt einmal mehr, unter welcher Belastung Athletinnen und Athleten stehen“, erklärte Moritz Anderten, Sportpsychologe am Olympiastützpunkt in Köln.
(L21/JUL.03123 Berliner Morgenpost, 29.07.2021, S. 22; Kampf gegen Dämonen)
- (20d) 13 Jahre stand auch der gemischte Chor Dickesbach unter Friedel Schmidts Leitung, ehe er von Maxim Schukov übernommen wurde.
(RHZ12/MAR.25654 Rhein-Zeitung, 21.03.2012, S. 21; Musikalisches Feuerwerk begeisterte Gäste)

Die Beispiele in (20) illustrieren, dass es Variation bezüglich des Vorkommens von Determinierern in FVGs gibt. Tabelle 6 fasst die Ergebnisse für die drei untersuchten FVG-Familien zusammen. Die relativen Häufigkeiten der Vorkommen der verschiedenen Determinierertypen ist relativ zur Gesamtheit determinierter NPs in der jeweiligen Familie.

Familie	Determinierer vorhanden						kein Determinierer vorhanden
	Gesamt	def. Artikel	indef. Artikel	Pronomen	GEN-NP	<i>kein</i>	
Prospektiv	991 (99,1%)	809 (81,63%)	160 (16,15%)	23 (2,32%)	0	0	9 (0,9%)
Vorgangspassiv	1028 (42,83%)	858 (83,46%)	41 (3,99%)	99 (9,63%)	28 (2,72%)	2 (0,19%)	1372 (57,17%)
Zustandspassiv	254 (16,93%)	202 (79,53%)	36 (14,17%)	16 (6,3%)	0	0	1246 (83,07%)

Tabelle 6

Vorkommen unterschiedlich determinierter Nomen in den FVGs der drei untersuchten Familien.

Die drei FVG-Familien variieren bezüglich der Häufigkeit, in der die PP-internen Nomen determiniert verwendet werden. Der größte Unterschied besteht dabei zwischen den FVGs der Prospektivfamilie, die mit 99% fast durchweg determinierte Nomen aufweisen, und den FVGs der Zustandspassivfamilie, deren PP-interne Nomen in lediglich 16,93% der Fälle determiniert sind. Die FVGs der Vorgangspassivfamilie liegen mit knapp 43% zwischen den beiden anderen Familien.

Innerhalb der einzelnen Familien stellt der definite Artikel den häufigsten Determinierertyp dar. Die Varianz zwischen den drei Familien ist dabei gering. Die relative Häufigkeit indefiniter Artikel ist in der Prospektiv- und der Zustandspassivfamilie etwa gleich hoch, die Vorgangspassivfamilie unterscheidet sich deutlich von den beiden anderen Familien.

Die Daten zeigen, dass sowohl Genitiv-NPs als auch die Negation mit *kein* nur in der Vorgangspassivfamilie und dort auch sehr marginalisiert vorkommen. Interessant ist dabei, dass diese Familie auch das höchste Vorkommen pronominaler Ausdrücke aufweist. Die Mehrheit der pronominalen Ausdrücke stellen Possessivpronomina dar. Da pronominale Formen und Genitiv-NPs in ihrem Auftreten beschränkt sind, konzentriere ich mich im weiteren Verlauf der Untersuchung auf den definiten und indefiniten Artikel.

Tabelle 6 gibt keine Informationen über das Vorkommen attributiver Modifikatoren innerhalb der untersuchten Familien. Auf Modifikatoren gehe ich in Kapitel 5.4 gesondert ein und habe sie daher aus Gründen der Übersichtlichkeit nicht in Tabelle 6 aufgenommen.

Im weiteren Verlauf dieses Kapitels möchte ich drei Fragen beantworten: (i) Wie homogen verhalten sich die FVGs der einzelnen Familien bezüglich der Artikelvarianz? Und (ii) werden die Artikel, wie in der Forschungsliteratur behauptet, enklitisch realisiert? Schließlich, (iii) unter welchen Bedingungen tritt der indefinite Artikel in den untersuchten FVGs auf?

5.2. Artikelvarianz innerhalb der einzelnen FVG-Familien

5.2.1. Prospektivfamilie

Tabelle 7 fasst das Vorkommen definiter und indefiniter Artikel innerhalb der FVGs der Prospektivfamilie zusammen. Die relative Häufigkeit bezieht sich nicht auf das absolute Vorkommen determinierter Nomen innerhalb der FVGs, sondern auf die Anzahl definiter, bzw. indefiniter Artikelverwendungen in Bezug auf die Gesamtheit an Artikelverwendungen.

In den meisten Fällen liegt der Anteil definiter Artikel bei $> 87\%$. Deutlich abweichend sind jedoch die beiden FVGs *vor dem Sieg stehen* und *vor einem Wechsel stehen*. Das erste FVG kommt nur in etwas mehr als der Hälfte aller Fälle mit definitivem Artikel vor, *vor einem Wechsel stehen* sogar nur in rund 11%. Dies ist auch das einzige FVG der Familie, das einen höheren Anteil indefiniter als definiter Verwendungen aufweist.

FVG	definitiver Artikel	indefinitiver Artikel	FVG	definitiver Artikel	indefinitiver Artikel
<i>Abschluss</i>	90 (94,74%)	5 (5,26%)	<i>Pleite</i>	96 (97,96%)	2 (2,04%)
<i>Bankrott</i>	97 (97%)	3 (3%)	<i>Ruin</i>	98 (100%)	0 (0%)
<i>Durchbruch</i>	83 (87,37%)	12 (12,63%)	<i>Sieg</i>	51 (54,84%)	42 (45,16%)
<i>Kollaps</i>	87 (95,6%)	4 (4,4%)	<i>Wechsel</i>	11 (11,11%)	88 (88,89%)
<i>Konkurs</i>	99 (99%)	1 (1%)	<i>Zusammenbruch</i>	97 (97%)	3 (3%)

Tabelle 7

Anteil definiter und indefiniter Artikel in der Prospektivfamilie.

Als einziges FVG ist *vor dem Ruin stehen* nicht mit dem indefiniten Artikel innerhalb der untersuchten Daten belegt. Dies erweist sich jedoch als ein Artefakt der Korpusdaten, da eine gezielte Suche nach einem Vorkommen mit Indefinitartikel den Treffer in (21) ergibt. Somit kann festgehalten werden, dass alle untersuchten FVGs der Familie auch mit dem Indefinitartikel vorkommen können.

- (21) Bettelaktionen, Rettungsversuche in letzter Minute, Defizite, Trainer- und Spielerentlassungen, Millionentransfers - alles emotionale Themen, nicht nur in Bern und Zürich, wo die Young Boys vor dem Ruin und die Grasshoppers vor einem Rekorddefizit stehen.

(A99/NOV.80666 St. Galler Tagblatt, 16.11.1999, Ressort: TB-SG (Abk.); Auf dem Teppich und damit im Espenmoos bleiben)

5.2.2. Vorgangspassivfamilie

In Tabelle 8 ist für die FVGs der Vorgangspassivfamilie aufgeführt, wie häufig diese mit definitivem, bzw. indefinitem Artikel vorkommen.

FVG	definitiver Artikel	indefinitiver Artikel	FVG	definitiver Artikel	indefinitiver Artikel
<i>Aufsicht</i>	29 (96,67%)	1 (3,33%)	<i>Herrschaft</i>	40 (97,56%)	1 (2,44%)
<i>Befehl</i>	72 (100%)	0	<i>Kommando</i>	60 (95,24%)	3 (4,76%)
<i>Belastung</i>	11 (31,43%)	25 (69,4%)	<i>Kontrolle</i>	45 (97,83%)	1 (2,17%)
<i>Beobachtung</i>	3 (100%)	0	<i>Kritik</i>	4 (100%)	0
<i>Besatzung</i>	1 (50%)	1 (50%)	<i>Leitung</i>	76 (100%)	0
<i>Beschuss</i>	3 (100%)	0	<i>Schirmherrschaft</i>	92 (100%)	0
<i>Betreuung</i>	5 (100%)	0	<i>Schutz</i>	41 (97,62%)	1 (2,38%)
<i>Bewachung</i>	4 (80%)	1 (20%)	<i>Strafe</i>	0	0
<i>Einwirkung</i>	46 (100%)	0	<i>Überwachung</i>	4 (100%)	0
<i>Federführung</i>	63 (100%)	0	<i>Verdacht</i>	43 (100%)	0
<i>Fittichen</i>	67 (100%)	0	<i>Vertrag</i>	0	0
<i>Fuchtel</i>	92 (100%)	0	<i>Verwaltung</i>	19 (82,61%)	4 (17,39%)
<i>Führung</i>	38 (92,68%)	3 (7,32%)			

Tabelle 8

Anteil definitiver und indefiniter Artikel in der Vorgangspassivfamilie.

Die FVGs der Vorgangspassivfamilie weisen eine Präferenz für den definiten gegenüber dem indefiniten Artikel auf. Drei Abweichungen von diesem Muster lassen sich in den Daten finden. Erstens, *vor der/einer Belastung stehen* kommt häufiger mit indefinitem als mit definitem Artikel vor. *Unter Besatzung stehen* kommt gleichhäufig mit beiden Artikeln vor, insgesamt überwiegt aber die Verwendung ohne Artikel (98% aller Fälle). Die FVGs *unter Strafe stehen* und *unter Vertrag stehen* kommen nur ohne Artikel vor. Fleischhauer (2021b) argumentiert bzgl. des FVGs *unter Vertrag stehen*, dass die Verwendung eines Artikels die konkrete Lesart von *Vertrag* evozieren würde, die mit *stehen* nur schwer aber mit *liegen* gut möglich ist (22).

(22) Der Personalausweis liegt unter dem Vertrag.
 (Fleischhauer 2021b, S. 355)

Ähnlich verhält es mit dem Nomen *Strafe*. Auch hier wird durch die Determination – An- vs. Abwesenheit eines Artikels – eine

Lesartenunterscheidung vorgenommen. Während *Strafe* in (23a) eine konkrete Interpretation hat und das Resultat des Strafens bezeichnet, hat das Nomen in (b) eine prozessuale Lesart und referiert auf den Prozess des Strafens (Fleischhauer 2022, S. 269).

(23a) Damit setze sich der Trend fort, bei Urteilen gegen Kriegs- und Zivildienstverweigerer unter einer Strafe zu bleiben, die nicht zu einem Eintrag ins Führungszeugnis führe.

(T01/JUN.26715 die tageszeitung, 06.06.2001, S. 19, Ressort: Berlin Aktuell; Totalverweigerer nicht vorbestraft)

(23b) Nach Auffassung der Wissenschaftler sollte Hilfe zur Selbsttötung grundsätzlich unter Strafe stehen.

(NUZ14/AUG.02383 Nürnberger Zeitung, 27.08.2014, S. 1; Experten legten Gesetzesvorschlag vor - Wird die ärztliche Sterbehilfe gelockert)

Die Nomen *Fittiche* und *Fuchtel* haben ursprünglich auch eine konkrete Bedeutung – ‚Flügel‘ und ‚kleines Schwert‘ –, diese scheinen aber weniger frequent als die übertragene Bedeutung ‚Kontrolle/ Schutz‘ zu sein. Es ist zu vermuten, dass nominale Determination dann möglich ist, wenn die konkrete Bedeutung (stark) verblasst ist (siehe Fleischhauer 2023b, S. 395ff. für eine Diskussion des FVGs *unter jemandes Fittichen stehen*). Auf diesen Aspekt komme ich im Zusammenhang mit der Diskussion des FVGs *unter Feuer stehen* in 5.2.3 zurück.

5.2.3. Zustandspassivfamilie

Die Auswertung der Vorkommenshäufigkeit des definiten und indefiniten Artikels innerhalb der FVGs der Zustandspassivfamilie ist in Tabelle 9 zusammengefasst.

Die FVGs der Zustandspassivfamilie stechen unter den drei untersuchten Familien bezüglich des Indefinitartikels hervor. Zwar kommt der indefinite Artikel lediglich in vier FVGs vor, der relative Anteil für die einzelnen FVGs ist jedoch deutlich höher als in den anderen Familien. In der absoluten Häufigkeit sind Artikel in dieser Familie – mit Ausnahme der beiden FVGs *unter Einfluss stehen* und *unter Zwang stehen* – selten. Der hohe relative Anteil des indefiniten Artikels ist durch die geringe Häufigkeit des definiten Artikels begünstigt (hier ist wiederum *unter Zwang stehen* eine Ausnahme).

FVG	definitiver Artikel	indefinitiver Artikel	FVG	definitiver Artikel	indefinitiver Artikel
<i>Alkohol</i>	0	0	<i>Medikamenten</i>	0	0
<i>Anklage</i>	14 (100%)	0	<i>Schock</i>	4 (100%)	0
<i>Anspannung</i>	17 (60,71%)	11 (39,29%)	<i>Spannung</i>	1 (14,29%)	6 (85,71%)
<i>Dampf</i>	0	0	<i>Stress</i>	10 (100%)	0
<i>Drogen</i>	0	0	<i>Strom</i>	0	0
<i>Druck</i>	9 (81,82%)	2 (18,18%)	<i>Zugzwang</i>	1 (100%)	0
<i>Einfluss</i>	81 (100%)	0	<i>Zwang</i>	60 (77,92%)	17 (22,08%)
<i>Feuer</i>	4 (100%)	0			

Tabelle 9
 Anteil definitiver und indefinitiver Artikel in der Zustandspassivfamilie.

Die Zustandspassivfamilie enthält FVGs mit konkreten PP-internen Nomen wie *Alkohol*, *Dampf*, *Drogen*, *Feuer*, *Medikamente* und *Strom*. Die fehlende Determination bei *Dampf* und *Strom* geht einher mit einer figurativen Interpretation der Nomen. *Feuer* ist ebenfalls figurativ gebraucht, kommt aber determiniert vor. Das DWDS¹¹ führt als eine Lesart von *Feuer* ‚innerer Schwung/ Begeisterung‘ auf. Dies deutet möglicherweise darauf hin, dass eine metaphorische Interpretation bereits stärker konventionalisiert ist. Das würde dazu passen, dass die Abwesenheit nominaler Determination nicht benötigt wird, um eine figurative Interpretation zu evozieren. Die Idee, dass Abwesenheit nominaler Determination eine Möglichkeit ist, eine figurative Interpretation herbeizuführen, benötigt weitere Untersuchung und soll an dieser Stelle nur als Hypothese verstanden werden.

Anders ist der Fall mit den Nomen *Alkohol*, *Drogen* und *Medikamente* gelagert. Diese Nomen werden wörtlich gebraucht; der Subjektreferent in (24) befindet sich unter dem Einfluss von Alkohol, bzw. wird von Alkohol z.B. in seinem Verhalten beeinflusst (vgl. Fleischhauer 2022, S. 272ff.; Fleischhauer 2023b, S. 397f.).

- (24) Der aggressive 42jährige stand bei seinem Manöver erheblich unter Alkohol.
 (M96/602.06735 Mannheimer Morgen, 12.02.1996; Eifersüchtiger Raser)

Die Nomen *Alkohol*, *Dampf*, *Drogen* und *Medikamente* kommen nur ohne Artikel vor. Hier liegt derselbe Fall wie bei dem Nomen *Vertrag* aus Kapitel 5.2.2. vor. Die Artikellosigkeit geht in diesem Fall auf die Referenzlosigkeit der Nomen zurück. Die Nomen haben lediglich eine generische Lesart, weisen aber keinen Bezug auf beispielsweise einen spezifischen Alkohol, bzw. eine spezifische Portion Alkohol auf. Es ist nicht überraschend, dass eine nicht-

¹¹ <https://www.dwds.de/wb/Feuer>; letzter Aufruf 06.03.2023.

referentielle Verwendung bei nicht metaphorisch interpretierten Konkreta stattfindet, womit eine wörtliche Interpretation des FVGs – unter einem bestimmten Alkohol (stehend) lokalisiert sein – ausgeschlossen ist.

5.3. Freie vs. enklitische Realisierung der Artikel

Im Deutschen können definite und indefinite Artikel entweder frei oder enklitisch, also in phonologisch reduzierter Form an eine vorhergehende Präposition klitisiert, realisiert werden (u. a. Nübling 1998, S. 275ff.). Während die Klitisierung des definiten Artikels sowohl im mündlichen als auch schriftlichen Sprachgebrauch zu finden ist, ist die Klitisierung des indefiniten Artikels eher auf den mündlichen Bereich – gesprochene Sprache und ihre Verschriftlichung – beschränkt (u.a. Schäfer, Sayatz 2014).

Für die untersuchten Präpositionen kommen die klitisierten Formen *unterm* (*unter* + *dem*), *untern* (*unter* + *den*) und *vorm* (*vor* + *dem*) vor. Eine Klitisierung des indefiniten Artikels ist für diese Präpositionen potentiell auch möglich, z.B. *vor'nem* in (5d) wiederholt als (25). Interessant an dem Beispiel ist, dass es die Klitisierung des indefiniten Artikels innerhalb eines FVGs der Prospektivfamilie zeigt.

(25) Mein Herz schlägt dermaßen schnell, dass ich kurz vor'nem Herzinfarkt stehe.¹²

Das Beispiel in (25) stammt nicht aus dem Deutschen Referenzkorpus. Alle aus dem Korpus extrahierte Beispiele kommen nur mit freiem Artikel vor. Innerhalb der untersuchten 5000 Korpusbelege gibt es somit kein einziges Vorkommen eines klitisierten Artikels. Dies ist erstaunlich, da in der Forschungsliteratur betont wird, dass die in FVGs vorkommenden Artikel immer klitisiert seien. Damit wird die Frage aufgeworfen, ob die untersuchten FVGs überhaupt enklitische Artikel erlauben.

Zur Beantwortung dieser Frage habe ich gezielt nach Vorkommen enklitischer Artikel innerhalb des FVG-Samples gesucht. Als Suchstring habe ich ‚&stehen /s0 {vorm/vorn/unterm} N‘ verwendet. Einige Nomen schieden aus der Untersuchung aus, da eine enklitische Verwendung des Artikels aus grammatischen Gründen nicht möglich ist, z.B. bei *unter der Aufsicht* (zu Restriktionen der Enklise, vgl. u.a. Nübling 1998). Die Untersuchung ist auf das Vorkommen des enklitischen definiten Artikels beschränkt, da der enklitische Indefinitartikel im DeReKo fast nicht vorkommt. Die Suche nach *vor'nem* ergab lediglich einen einzigen Treffer im Archiv W des DeReKo. Dass der enklitische Indefinitartikel potentiell zumindest in den FVGs der Prospektivfamilie vorkommen kann, zeigt aber bereits Beispiel (25).

¹² <https://context.reverso.net/%C3%BCbersetzung/deutsch-englisch/kurz+vor+%27nem>; Aufruf: 27.02.2023.

Tabelle 10 fasst die Ergebnisse der Untersuchung sortiert nach FVG-Familien zusammen. ‚N‘ gibt die Gesamtanzahl der Vorkommen des klitisierten Artikels innerhalb des jeweiligen FVGs an. Alle FVGs der Prospektivfamilie, die eine Klitisierung des Artikels erlauben, kommen auch mit klitisiertem Artikel vor. In den anderen beiden Familien ist der Gebrauch des enklitischen Artikels deutlich eingeschränkter. Einige der FVGs, die nicht mit enklitischem Artikel belegt sind, kommen in Untersuchungssample mit freiem definiten Artikel vor, z.B. *unter Befehl stehen*, *unter jemandes Fittichen stehen*, *unter Feuer stehen*, *unter Stress stehen*, *unter Zugzwang stehen*.

FVG (Prospektiv)	N	FVG (Vorgangspassiv)	N	FVG (Zustandspassiv)	N
<i>vom Abschluss</i>	85	<i>unterm Befehl</i>	0	<i>unterm Alkohol</i>	0
<i>vom Bankrott</i>	21	<i>unterm Beschuss</i>	0	<i>unterm Dampf</i>	0
<i>vom Durchbruch</i>	14	<i>untern Fittichen</i>	0	<i>unterm Druck</i>	3
<i>vom Kollaps</i>	47	<i>unterm Kommando</i>	1	<i>unterm Einfluss</i>	2
<i>vom Konkurs</i>	29	<i>unterm Schutz</i>	6	<i>unterm Feuer</i>	0
<i>vom Ruin</i>	11	<i>unterm Verdacht</i>	1	<i>unterm Schock</i>	1
<i>vom Sieg</i>	11			<i>unterm Stress</i>	0
<i>vom Wechsel</i>	43			<i>unterm Strom</i>	1
<i>vom Zusammenbruch</i>	23			<i>unterm Zugzwang</i>	0
				<i>unterm Zwang</i>	1

Tabelle 10
 Absolute Häufigkeit des Vorkommens enklitischer definitiver Artikel innerhalb der untersuchten FVG-Familien.

In der Forschungsliteratur wird angeführt, dass innerhalb von Präpositionalphrasen klitisierte Artikel eine größere Vorkommenshäufigkeit als freie Artikel aufweisen, bzw. freie Artikel gar nicht vorkommen sollen (vgl. Kapitel 3).¹³ In ihrer Untersuchung des Artikelenklise merkt Cieschinger (2016) an, dass in den von ihr untersuchten Korpora die klitisierten Formen in mehr als 90% aller Fälle verwendet werden und die Sequenz aus Präposition und freiem Artikel entsprechend in weniger als 10% aller Fälle. Diese Ergebnisse betreffen generelle die Kombination von ‚Präposition + Artikel‘ und sind nicht auf Funktionsverbgefüge bezogen. Für die untersuchten FVGs kann diese Tendenz nicht bestätigt werden. Es muss jedoch beachtet werden, dass die absoluten Häufigkeiten in Tabelle 10 nicht mit denen aus den Tabellen 7 bis 9 verglichen werden können. Tabelle 10 spiegelt die absolute Häufigkeit des Vorkommens klitisierte definitiver Artikel innerhalb der untersuchten FVGs im kompletten Archiv W wider. Die absolute Häufigkeit in den Tabellen 7 bis

¹³ Die Angaben beziehen sich auf die kontrahierten Formen *im*, *am*, *zum*, *zur*, *vom*, *beim*, die die größte schriftsprachliche Akzeptanz aufzuweisen scheinen (vgl. Cieschinger 2016, S. 5).

9 bezieht sich dagegen nur auf den untersuchten Ausschnitt aus 100 Zufallsbelegen pro Funktionsverbgefüge.

Die in dieser Arbeit untersuchten Daten sind auch nur begrenzt mit Cieschingers Daten vergleichbar, da *unterm* und *vorm* in ihrer Untersuchung nicht vorkommen. Eine Recherche im DeReko (Archiv W) nach den Suchstrings ‚vorm‘, ‚vor dem‘, ‚unterm‘ und ‚unter dem‘ zeigt, dass die entsprechenden Sequenzen aus Präposition und freiem Artikel im Korpus deutlich häufiger vorkommen als die klitisierten Formen (vgl. Tabelle 11). Die Ergebnisse, die in Tabelle 8 zu sehen sind, sind exakt spiegelbildlich zu den von Cieschinger untersuchten Konstruktionen. Die geringere allgemeine Vorkommenshäufigkeit der klitisierten Formen *unterm* und *vorm* erklärt auch, warum diese in der in Abschnitt 5.1 beschriebenen Korpusuntersuchung nicht vorkamen.

Form	Häufigkeit	Form	Häufigkeit
<i>vorm</i>	98443 (2,71%)	<i>unterm</i>	144427 (7,25%)
<i>vor dem</i>	3538538 (97,29%)	<i>unter dem</i>	1847296 (92,25%)
Gesamt	3636981 (100%)	Gesamt	1991723 (100%)

Tabelle 11

Absolute Häufigkeit klitisierter Artikel im Vergleich zu Sequenzen aus Präposition und freiem Artikel.

5.4. Indefiniter Artikel

Die Ergebnisse der Korpusuntersuchung aus Abschnitt 5.1 haben gezeigt, dass in allen FVG-Familien auch der indefinite Artikel vorkommt. Er kommt in keiner Familie häufiger als der definite Artikel vor, wobei dies nicht für die einzelnen FVGs der Familien gelten muss. Entgegen dem eindeutigen Trend zur Verwendung des definiten Artikels kommt beispielsweise das prospektive FVG *vor einem Wechsel stehen* in 89% aller determinierten Fälle mit dem indefiniten Artikel vor und nur in rund 11% der Fälle mit dem definiten Artikel.

Die einzelnen FVG-Familien unterscheiden sich hinsichtlich der Bedingungen, unter denen der indefinite Artikel verwendet wird. In den FVGs der Vorgangs- und Zustandspassivfamilie kommt der indefinite Artikel primär in Kombination mit einem Attribut vor. Die Auswertung in Tabelle 12 zeigt, dass in der Vorgangspassivfamilie der indefinite Artikel fast durchgängig mit einem attributiven Adjektiv auftritt.

	Vorgangspassivfamilie	Zustandspassivfamilie	Prospektivfamilie
Indefiniter Artikel mit Attribut	39 (95,12%)	21 (58,33%)	37 (23,13%)
Indefiniter Artikel ohne Attribut	2 (4,88%)	15 (41,67%)	123 (76,87%)
Gesamt	41	36	160

Tabelle 12
 Vorkommenshäufigkeit indefiniter Artikel mit und ohne Attribut.

In der Zustandspassivfamilie tritt der indefinite Artikel in fast 60% der Fälle in Kombination mit einem attributiven Adjektiv auf. Tabelle 13 fasst die Ergebnisse für die einzelnen FVGs der Zustandspassivfamilie, die überhaupt mit dem indefiniten Artikel vorkommen, zusammen.

FVG	indefiniten Artikel	Attribut + indefiniten Artikel
<i>Anspannung</i>	11	9 (81,82%)
<i>Druck</i>	2	2 (100%)
<i>Spannung</i>	6	2 (33,33%)
<i>Zwang</i>	17	8 (47,06%)
Gesamt	36	21 (58,33%)

Tabelle 13
 Vorkommenshäufigkeit des indefiniten Artikels mit und ohne Attribut in FVGs der Zustandspassivfamilie.

Interessant ist insbesondere das FVG *unter Spannung stehen*. In allen vier Fällen, in denen der indefinite Artikel ohne begleitenden attributiven Modifikator verwendet wird, hat *Spannung* eine physikalische Bedeutung, etwa ‚elektrische Spannung‘ (26a). In den anderen beiden Fällen, in denen der Artikel mit Attribut vorkommt, hat *Spannung* die Bedeutung ‚mechanische Spannung‘, wie in (26b). Alle Beispiele analog zu (26a) weisen eine Possessivkonstruktion der Form *von X Volt* auf. Funktional erbringt die *von-PP* dieselbe Leistung wie ein attributiver Modifikator in der Lizenzierung des indefiniten Artikels.

(26a) Davor habe man allerdings, so Schneider, die Anlage sichern müssen, da sie unter einer Spannung von 110000 Volt stehe.

(A17/MAI.10585 St. Galler Tagblatt, 24.05.2017, Ressort: Ostschweiz; Kalter Morgenkaffee für die Ostschweiz)

(26b) Leute, die mit Kettensägen gefällte Bäume zersägten, wussten nicht, dass frisches Fallholz oft unter einer Spannung steht und beim Zersägen dann wie eine Feder reagiert.

(E96/SEP.21558 Tages-Anzeiger, 09.09.1996, S. 12, Ressort: Kehrseite; Nach Sturm Wassersnot)

Im Zusammenhang mit dem FVG *unter Zwang stehen* lassen sich zwei weitere Strategien einer NP-externen Modifikation des nominalen Elements beobachten. Die Art des Zwangs kann durch einen *zu*-Infinitiv (event. auch in nominalisierter Form) (27a) oder durch einen Relativsatz (27b) spezifiziert werden.

(27a) Grundsätzlich stehen alle Banken unter einem Zwang zu wachsen.
(A17/DEZ.10199 St. Galler Tagblatt, 28.12.2017, Ressort: Wirtschaft; „Die Neuen müssen sich erst beweisen“)

(27b) Er habe geradezu unter einem Zwang gestanden, den Auschwitz-Prozess besuchen zu müssen, sagte der amerikanische Dramatiker Arthur Miller („Tod eines Handlungsreisenden“), 49, Ende Februar dieses Jahres in Frankfurt.
(S64/DEZ.00148 Der Spiegel, 09.12.1964, S. 145; Jedermanns Juden)

Insgesamt gibt es in den Zustandspassivdaten damit nur drei Belege, in denen ein indefiniter Artikel nicht mit einem NP-internen attributiven Modifikator oder einer NP-externen Modifikation vorkommt. Wenn die NP-externe Modifikation mit einbezogen wird, dann kommen in insgesamt 91,6% der Fälle der indefinite Artikel mit einem NP-internen oder externen Attribut vor.

Ein anderes Bild bietet die Prospektivfamilie. In nur rund 23% alle Fälle kommt der indefinite Artikel mit einem Attribut vor. Die Auswertung für die einzelnen FVGs der Familie in Tabelle 14 zeigt, dass lediglich *vor einem Sieg stehen* eine stärkere Präferenz für die Kombination von indefinitem Artikel und attributivem Adjektiv aufweist. In den Daten gibt es auch keinen signifikanten Anstieg von Attributen, wenn NP-externe Modifikatoren hinzugezählt werden.

FVG	indefiniten Artikel	Attribut + indefiniten Artikel
<i>Abschluss</i>	5	1 (20%)
<i>Bankrott</i>	3	1 (33,33%)
<i>Durchbruch</i>	12	2 (16,67%)
<i>Kollaps</i>	4	0 (0%)
<i>Konkurs</i>	1	0 (0%)
<i>Pleite</i>	2	1 (50%)
<i>Sieg</i>	42	28 (66,67%)
<i>Wechsel</i>	88	4 (4,55%)
<i>Zusammenbruch</i>	3	0 (0%)
Gesamt	160	37 (23,13%)

Tabelle 14

Vorkommenshäufigkeit des indefiniten Artikels mit und ohne Attribut in FVGs der Prospektivfamilie.

Wir können also einen deutlichen Unterschied in der Lizenzierung des indefiniten Artikels zwischen den beiden Passivfamilien und der Prospektivfamilie erkennen. Für eine zukünftige Untersuchung stellt sich die Frage, wieso diese Unterschiede zwischen den FVG-Familien bestehen.

6. Auswertung der Ergebnisse

Die Untersuchung der drei FVG-Familien hat ergeben, dass Funktionsverbgefüge eine größere Artikelvarianz aufweisen als oftmals in der Forschungsliteratur behauptet. Während die Forschungsliteratur Artikelvarianz auf die Opposition zwischen enklitischem definitem Artikel und keinem Artikel einschränkt, zeigen die Daten, dass das volle Spektrum an Determinierern – freier definiten und indefiniten Artikel, enklitischer definiten und indefiniten Artikel, Demonstrativa und Pronomina – zu finden ist. Somit ist festzuhalten, dass die FVGs der untersuchten Familien keine Beschränkungen in der Artikelvarianz aufweisen.

Die Daten haben zugleich gezeigt, dass die einzelnen FVG-Familien unterschiedliche Präferenzen bezüglich der Artikelvarianz aufweisen. FVGs der Prospektivfamilie kommen nur in 3% aller Fälle ohne Artikel vor, wohingegen artikellose Vorkommen innerhalb der beiden Passivfamilien deutlich frequenter sind (~60% in der Vorgangspassivfamilie und ~83% in der Zustandspassivfamilie). Aber auch innerhalb der Familien gibt es Varianz, wie die Auswertungen in Kapitel 5.2 gezeigt haben. Daraus lässt sich schlussfolgern, dass Generalisierungen über Funktionsverbgefüge hinweg, wie sie in der in Kapitel 3 diskutierten Forschungsliteratur erfolgen, weder möglich noch sinnvoll sind. Funktionsverbgefüge weisen kein homogenes Verhalten bezüglich Artikelvarianz auf, auch wenn sich systematische Muster finden lassen. Diese Muster sind aber auf der Ebene der FVG-Familien angesiedelt und können somit nicht als allgemeine Eigenschaften von Funktionsverbgefügen angesehen werden. Dieses Resultat ist kompatibel mit Harms (2021) Beobachtungen, dass der Artikelgebrauch in FVGs u.a. vom semantischen Typ des FVG-internen Nomens abhängt.

In einer der Fragestellung aber nicht im Umfang vergleichbaren Untersuchung kommt Storrer (2006, S. 292) zu dem Ergebnis, dass „die Beschränkung der Determination ein charakteristisches Merkmal“ von FVGs sei. Storrers korpuslinguistische Untersuchung ist auf die beiden FVGs *in Verbindung treten* und *in Kontakt treten* beschränkt. Diese kommen in ihren Daten fast vollständig ohne Artikel vor, lediglich *in Verbindung treten* tritt einmal (0,5% aller Fälle) mit indefinitem Artikel auf (Storrer 2006, S. 287). Eine Untersuchung der Artikelvarianz im Archiv W des Deutschen Referenzkorpus mittels der in Kapitel 5 vorgestellten Methodik erbrachte Ergebnisse, die nahezu identisch mit denen Storrers sind. Die Ergebnisse sind in Tabelle 15 zusammengefasst. Lediglich das FVG *in Verbindung treten* kommt mit PP-internem Artikel vor, in beiden Fällen tritt der indefinite Artikel in Kombination mit einem attributiv gebrauchten Adjektiv auf.

	<i>in Kontakt treten</i>	<i>in Verbindung treten</i>
FVG-internes Nomen ohne Artikel	100 (100%)	98 (98%)
FVG-internes Nomen mit Artikel	0	2 (2%)
Gesamt	100	100

Tabelle 15

Häufigkeit des Vorkommens der FVGs in Kontakt treten und in Verbindung treten mit und ohne Artikel.

Aus diesen Ergebnissen zu schlussfolgern, dass die FVGs eine Beschränkung bezüglich der in ihnen vorkommenden Artikel aufweisen würden, wäre jedoch voreilig. Eine gezielte Suche nach dem Vorkommen des definiten, bzw. indefiniten Artikels innerhalb der beiden FVGs zeigt, dass solche Belege existieren. Mittels der Suchanfrage ‚&treten /s0 in {den/einen} /w2 Kontakt‘ konnten 60 Belege für das Vorkommen des definiten Artikels (28a) und 66 Belege für das Vorkommen des indefiniten Artikels (28b) innerhalb des FVGs gefunden werden. Durch ihre geringe Anzahl gehen sie lediglich in der Gesamtheit der Treffer unter. Die Suchanfrage ‚&treten /s0 in {die/eine} /w2 Verbindung‘ erbrachte einen Treffer für die Verwendung des definiten Artikels und 29 für die Verwendung des indefiniten Artikels innerhalb des FVGs *in Verbindung treten*. Die geringe Anzahl an Vorkommen definitiver Artikel mag im Fall dieses FVGs tatsächlich auf eine Beschränkung der Artikelvarianz hindeuten, deren Ursache jedoch noch genauer zu untersuchen ist.

- (28a) Mit «HSG Hautnah» möchte die Universität St. Gallen in den direkten Kontakt mit der Bevölkerung treten.
(A13/APR.01683 St. Galler Tagblatt, 05.04.2013, S. 47; Vom Hügel in die Arena steigen)
- (28b) Schüler aus Ostvorpommern und der italienischen Provinz Reggio Calabria sind in einen kreativen Kontakt getreten, um etwas für demokratische Kultur und gegen diskriminierende Ausgrenzungen in der Gesellschaft zu bewirken.
(NKU09/OKT.04742 Nordkurier, 14.10.2009; Aus feucht-fröhlicher Tour wird ein schicksalshafter Tag)
- (28c) [...] so treten wir durch den Genuss der eucharistischen Gaben ein in die engste Verbindung mit Jesus Christus.
(RHZ15/JUN.02570 Rhein-Zeitung, 03.06.2015, S. 14; Ein Fest unseres Glaubens)
- (28d) Ein Konzert, in dem Improvisation und Komposition in eine neue Verbindung treten: [...].
(E10/JUL.01326 Tages-Anzeiger, 09.07.2010, S. 27; Wenn heute sich mit gestern vermischt...)

Die Ergebnisse der Untersuchung zeigen einerseits, dass ein daten-orientierter *bottom-up* Ansatz zur Untersuchung von Funktionsverbgefügen adäquater als der insbesondere in der älteren Forschungsliteratur verwendete *top-down* Ansatz ist. Andererseits zeigt die Diskussion der Ergebnisse von Storrers Untersuchung, dass auf Grundlage einer kleinen Auswahl korpuslinguistischer

Belege Verallgemeinerungen nur mit Vorsicht zu machen sind. Wie die Diskussion um die beiden *treten*-FVGs zeigt, kommen manche FVGs nur selten mit PP-internem Artikel vor, woraus aber nicht auf eine generelle Beschränkung, sondern allenfalls auf eine Präferenz bezüglich der Artikel geschlossen werden kann.

7. Fazit

Ausgangspunkt der vorliegenden Arbeit war die in der Forschungsliteratur seit längerem wiederholt auftretende Behauptung, dass Funktionsverbgefüge nur eine eingeschränkte Artikelvarianz – Opposition zwischen enklitischem definitem Artikel und keinem Artikel – aufweisen würden. Auf Grundlage einer korpusbasierten Untersuchung dreier distinkter FVG-Familien, die mit dem Funktionsverb *stehen* gebildet werden, konnte gezeigt werden, dass die untersuchten FVGs keine generellen Einschränkungen bezüglich der Artikelvarianz aufweisen. Einschränkungen bestehen lediglich auf Ebene individueller FVGs, dies betrifft vor allem FVGs mit Konkreta in PP-interner Position (z.B. *unter Vertrag stehen*, *unter Alkohol stehen*).

FVGs weisen somit keine generelle Beschränkung der Artikelvarianz auf, obgleich einzelne FVG-Familien und innerhalb der Familien die individuellen FVGs unterschiedliche Präferenzen bzgl. der Artikelwahl zeigen. Dies konnte auch noch einmal anhand der beiden FVGs *in Kontakt treten* und *in Verbindung treten* gezeigt werden. Diese FVGs weisen, wie auch Storrers Untersuchung nahelegt, eine Präferenz für ein artikelloses Vorkommens des PP-internen Nomens auf. Aber, wie meine Untersuchung im letzten Kapitel gezeigt hat, dies ist nur eine Präferenz und keine absolute Beschränkung. Belege der beiden FVGs mit determiniertem Nomen sind zwar selten, aber sind in natürlichen Sprachdaten zu finden.

Eingangs hatte ich auf von Polenz Unterscheidung zwischen Funktionsverbgefügen und Nominalisierungsverbgefügen (z.B. *eine Frage stellen*) hingewiesen. Wie in Kapitel 3 gezeigt, weisen die Nominalisierungsverbgefüge keine Beschränkung in der Artikelvarianz auf. Daraus leitete ich die Frage ab, ob sich FVGs und Nominalisierungsverbgefüge bezüglich Beschränkungen in der Artikelvarianz unterscheiden würden. Die Ergebnisse der vorliegenden Untersuchung legen nahe, dass dies nicht so ist. Beide Konstruktionstypen weisen Artikelvarianz auf, die Faktoren, die diese beeinflussen, unterscheiden sich aber möglicherweise. Dies muss jedoch in einer Folgestudie, die ein größeres Repertoire an Nominalisierungsverbgefügen berücksichtigt, untersucht werden.

Ein wesentliches Resultat der vorliegenden Untersuchung ist, dass Restriktionen bezüglich der Artikelvarianz kein Kriterium zur Identifikation von FVGs oder zur Abgrenzung der FVGs von anderen Typen prädikativer

Konstruktionen darstellen. Eine übereinstimmende Schlussfolgerung, jedoch basierend auf einer anderen empirischen Grundlage, zieht Kambers (2006, S. 29) in seiner Monographie zu Funktionsverbgefügen.

Die vorliegende quantitative Untersuchung stellt eine Grundlage für eine darauf aufbauende qualitative Studie dar, in der die semantischen Faktoren, die die Artikelverwendung steuern, konkreter untersucht werden müssten. Als Startpunkt für eine solche Untersuchung bietet sich insbesondere der Unterschied zwischen den beiden Passivfamilien und der Prospektivfamilie bezüglich der Lizenzierung des indefiniten Artikels an.

Bionote: Jens Fleischhauer finished his dissertation and habilitation in general linguistics at Heinrich-Heine-University Düsseldorf. He is primarily working on verb semantics and the syntax-semantics interface from a cross-linguistic perspective. At Heinrich-Heine-University, he headed a funded project on the composition of light verb constructions. In addition, he is one of the series editors of the newly established book series ‘Studies in Role and Reference Grammar’ published at Mouton.

Author’s address: fleischhauer@phil.uni-duesseldorf.de

Acknowledgements: The research reported in this paper is funded by the German Research Foundation as part of the project “Funktionsverbgefüge: Familien & Komposition” (HE 8721/1-1).

Literatur

- Boogards M. and Fleischhauer J. 2023, *Prospective aspect constructions in West Germanic. A comparative corpus study of German and Dutch*, in “Leuvense Bijdragen – Leuven Contribution in Linguistics and Philology” 104, pp. 5-37.
- Bresson D. 1988, *Der nominale Teil in den Funktionsverbgefügen*, in “Cahiers d’études germaniques” 14, pp. 111-122.
- Cieschinger M. 2016, *The Contraction of Preposition and Definite Article in German Semantic and Pragmatic Constraints*. Dissertation: Universität Osnabrück.
- Davidson D. 1980, *The individuation of events*, in Davidson D. (Hrsg.), *Essays on Action and Events*, Clarendon, Oxford, pp. 163-180.
- Davidson D. 1985, *Die logische Form der Handlungssätze*, in Davidson D. (Hrsg.), *Handlung und Ereignis*, Suhrkamp, Frankfurt am Main, pp. 155-177.
- Eisenberg P. 2006, *Funktionsverbgefüge – Über das Verhältnis von Unsinn und Methode*, in Breindl E., Gunkel L. und Strecker B. (Hrsg.), *Grammatische Untersuchungen*, Narr, Tübingen, pp. 297-318.
- Eisenberg P. 2013, *Grundriss der deutschen Grammatik. Der Satz*. J.B. Metzler, Stuttgart/Weinar, 4. Auflage.
- Engelen B. 1968, *Zum System der Funktionsverbgefüge*, in „Wirkendes Wort“ 18, pp. 289-303.
- Fleischhauer J. 2023a, *Prospective aspect and current relevance – A case study of the German prospective stehen vor NP construction*, in “Journal of Germanic linguistics” 35 [4], pp. 371-408.
- Fleischhauer J. 2023b, *German light verb constructions with non-eventive nouns*, in Colson J.-P. (Hrsg.), *Phraseology, constructions and translation*, Presses Universitaires de Louvaine, Louvain-la-Neuve, pp. 387-398.
- Fleischhauer J. 2022, *Stehen unter-Funktionsverbgefüge und ihre Familien*, in „Zeitschrift für germanistische Linguistik“ 50 [2], pp. 247-288.
- Fleischhauer J. 2021a, *Families of light verb constructions – A corpus study on German stehen unter-LVCs*, in “Proceedings of 17th Workshop on Multiword Expressions (MWE 2021)”, Association for Computational Linguistics, Online, pp. 63-69. [doi: [10.18653/v1/2021.mwe-1.8](https://doi.org/10.18653/v1/2021.mwe-1.8)]
- Fleischhauer J. 2021b, *Warum steht der Fußballspieler unter Vertrag? – Eine Fallstudie zu Funktionsverbgefügen des Typs ‘stehen unter’ NP*, in „Sprachwissenschaft“ 46 [3], pp. 343-375.
- Fleischhauer J. and Gamerschlag Th. 2019, *Deriving the meaning of light verb constructions – a frame account of German stehen 'stand'*, in Juchem-Grundmann C., Pleyer Michael and Pleyer Monica (Hrsg.), *Yearbook of the German Cognitive Linguistics Association, Vol. 7*, Mouton de Gruyter, Berlin/Boston, pp. 137-156. [doi: [10.1515/gcla-2019-0009](https://doi.org/10.1515/gcla-2019-0009)]
- Fleischhauer J., Gamerschlag Th., Kallmeyer L. and Petitjean S. 2019, *Towards a compositional analysis of German light verb constructions (LVCs) combining Lexicalized Tree Adjoining Grammar (LTAG) with frame semantics*. in “Proceedings of the 13th International Conference on Computational Semantics - Long Papers”, Association for Computational Linguistics, Gothenburg, pp. 79-90.
- Fleischhauer, J. and Turus D. 2021, *Der Angeklagte steht unter Schutz, wird er aber auch geschützt? – Eine Analyse passivischer Funktionsverbgefüge des Typs stehen unter*, in Jelitto-Piechulik, G., Jokieli, M., Książek and Pelka, D. (Hrsg.), *Germanistische Werkstatt 11*, Uniwersytet Opolski, Opole, pp. 73-84.

- Fleischhauer J. and Turus D. 2022, *Families of Light Verb Constructions in German Legal Language Contexts: A Case Study on the 'Passive'-Family of stehen unter-Light Verb Constructions*, in Green D. and Green L. (Hrsg.), *Contemporary Approaches to Legal Linguistics*, LIT Verlag, Münster, pp. 173-198.
- Fuhrhop N. 2007, *Nominale: flektiert und unflektiert*, in „Zeitschrift für germanistische Sprachwissenschaft“ 34 [3], pp. 267-285.
- Gamerschlag Th., Petersen W. and Ströbel L. 2013, *Sitting, standing, and lying in frames: A frame-based approach to posture verbs*, in Bezhanishvili G., Löbner S., Marra V. and Richter F. (Hrsg.), *Selected Papers of the 9th International Tbilisi Symposium on Logic, Language, and Computation*, Springer, Berlin, pp. 73-93.
- Harm V. 2021, *Funktionsverbgefüge des Deutschen*, De Gruyter, Berlin.
- Heine A. 2020, *Zwischen Grammatik und Lexik. Ein forschungsgeschichtlicher Blick auf Funktionsverbgefüge*, in de Knop S. und Hermann M. (Hrsg.), *Funktionsverbgefüge im Fokus*, De Gruyter, Berlin, pp. 15-38.
- Helbig G. 1984, *Probleme der Beschreibung von Funktionsverbgefügen im Deutschen*, in Helbig G. (Hrsg.), *Studien zur deutschen Syntax, Band 2*, Verlag Enzyklopädie Leipzig, Leipzig, pp. 163-188.
- Herrlitz W. 1973, *Funktionsverbgefüge vom Typ 'in Erfahrung bringen'*, Niemeyer, Tübingen.
- Kamber A. 2008, *Funktionsverbgefüge – empirisch*, Niemeyer, Tübingen.
- Kaufmann I. 1995, *Konzeptuelle Grundlagen semantischer Dekompositionsstrukturen*, Niemeyer, Tübingen.
- Krifka M. 1991, *Massennomina*, in von Stechow A. und Wunderlich D. (Hrsg.), *Semantik – Ein internationales Handbuch zeitgenössischer Forschung*, De Gruyter, Berlin, pp. 399-417.
- Leibniz-Institut für Deutsche Sprache. 2020. COSMAS II (Corpus Search, Management and Analysis System). Mannheim: Leibniz-Institut für Deutsche Sprache.
- Leibniz-Institut für Deutsche Sprache. 2021. Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2021-I (Release vom 02.02.2021). Mannheim: Leibniz-Institut für Deutsche Sprache.
- Leiss E. 2000, *Artikel und Aspekt*. Walter de Gruyter, Berlin/ New York.
- Maienborn C. 2011, *Event Semantics*, in Maienborn C., von Heusinger K. and Portner P. (Hrsg.), *Semantics. An International Handbook of Natural Language Meaning, Vol. 1*, de Gruyter, Berlin, pp. 802-829.
- Nübling D. 1998, *Wann werden die deutschen Präpositionen flektieren? Grammatisierungswege zur Flexion*, in Fabri R., Ortman A. and Parodi T. (Hrsg.), *Models of Inflection*, Niemeyer, Tübingen, pp. 266-289.
- Parsons T. 1990, *Events in the semantics of English*, MIT Press, Cambridge, MA.
- Reich I. 2017, *On the omission of articles and copulae in German newspaper headlines*, in “Linguistic Variation” 17 [2], pp. 186-204. [doi 10.1075/lv.14017.rei]
- Sandig B. 1971, *Syntaktische Typologie der Schlagzeile. Möglichkeiten und Grenzen der Sprachökonomie im Zeitungsdeutsch*, Dissertation, Universität zu Heidelberg.
- Schäfer R. und Sayatz U. 2014, *Die Kurzformen des Indefinitartikels im Deutschen*, in “Zeitschrift für germanistische Sprachwissenschaft” 33 [2], pp. 215-250.
- Smirnova E. and Stöber V. 2022, *Verbo-Nominal Constructions with kommen 'come' in German*, in “Frames and Constructions” 14 [1], pp. 121-149.
- Storrer A. 2006, *Zum Status der nominalen Komponenten in Funktionsverbgefügen*, in Breindl E., Gunkel L. und Strecker B. (Hrsg.), *Grammatische Untersuchungen*, Narr, Tübingen, pp. 275-296.

- van Pottelberge J. 2001, *Verbonominale Konstruktionen, Funktionsverbgefüge*, Universitätsverlag C. Winter, Heidelberg.
- Vater H. 1979, *Das System der Artikelformen im gegenwärtigen Deutsch*, Niemeyer, Tübingen.
- von Polenz P. 1963, *Funktionsverben im heutigen Deutsch. Sprache in der rationalisierten Welt*, Schwann, Düsseldorf.
- von Polenz P. 1987, *Funktionsverben, Funktionsverbgefüge und Verwandtes. Vorschläge zur satzsemantischen Lexikographie*, in "Zeitschrift für germanistische Linguistik" 15, pp. 169-189.
- Wunderlich D. und Herweg M. 1991, *Lokale & Direktionale*, in von Stechow A. und Wunderlich D. (Hrsg.), *Semantik – Ein internationales Handbuch zeitgenössischer Forschung*, De Gruyter, Berlin, pp. 758-785
- Zifonun G., Hoffmann L. und Strecker B. 1997, *Grammatik der deutschen Sprache. Band 1 & 2*, De Gruyter, Berlin.

ENTRE TERMINOLOGIE, TERMINOGRAPHIE ET HUMANITÉS NUMÉRIQUES

Le projet DIACOM-fr

PAOLO FRASSI
UNIVERSITÀ DEGLI STUDI DI VERONA

Abstract – The large-scale *Digital Humanities applied to foreign languages and literature* project (2018-2022) has enabled the Department of Foreign Languages and Literatures of the University of Verona to apply a large part of its linguistic, literary and didactic research to the digital humanities. The DIACOM-fr project (terminological database concerning the domain of international trade), which is part of the linguistic axis of this project, has several objectives: 1) to compile a corpus in the field of international trade (and make it available online for public consultation); 2) to develop a digital tool for interrogating this corpus; 3) to extract terminological units; 4) to compile a digital terminology database inspired by Explanatory and Combinatorial Lexicology. The purpose of this contribution is to present the objectives of the project, which has now come to an end.

Keywords: terminology; terminography; international trade; terminological database; terminological networks.

1. Introduction

La constitution d'une base de données terminologiques présente des enjeux théoriques et méthodologiques majeurs, notamment l'identification des entités terminologiques (simples et complexes), leur affichage, la prise en compte de la synchronie et/ou de la diachronie, la possibilité d'exploiter la base à des fins didactiques ou à d'autres fins, etc. Les décisions prises à ce niveau se reversent sur les modèles informatiques exploités pour la constitution de la base elle-même.

Nous illustrons, dans les lignes qui suivent, l'expérience du projet DIACOM-fr ; il s'agit d'un projet original, car il représente l'une des rares bases de données terminologiques inspirées de la Lexicologie Explicative et Combinatoire (Mel'čuk, Clas, Polguère 1995), affichant les entités terminologiques et leurs propriétés sous forme de *réseau terminologique* et valorisant, en plus de leurs caractéristiques sémantiques et des liens syntagmatiques et paradigmatiques des termes, leurs propriétés diachroniques et leurs propriétés didactiques.

Le projet s'est achevé en décembre 2022 et s'est échelonné sur plusieurs étapes que nous allons énumérer par la suite :

1. La création d'un corpus concernant le domaine du commerce international ;
2. L'exploitation d'un logiciel permettant la consultation du corpus ;
3. L'extraction automatique de termes simples et complexes ;
4. Le classement des termes complexes ;
5. L'identification des propriétés des termes ;
6. La représentation des entités terminologiques et de leurs propriétés dans une base de données terminologiques de type *réseau terminologique*.

Dans cette contribution nous nous proposons de présenter l'essentiel de chacune de ces étapes.

2. La constitution du corpus

La constitution d'un corpus représente souvent le point de départ d'un projet de type terminologique ; le fonctionnement des entités terminologiques en contexte est essentiel pour les réflexions théoriques concernant les propriétés des termes ainsi que pour l'activité terminographique.

Tout corpus terminologique se doit d'être représentatif du domaine en question – dans notre cas il s'agit du domaine du commerce international –, d'où l'adoption de deux critères, notamment le critère thématique et le critère textuel. S'ajoute un troisième critère, le critère chronologique, qui fait la spécificité de notre approche de type diachronique.

Par ce dernier critère – le critère chronologique –, nous avons voulu créer un corpus témoignant de l'évolution du domaine en question à travers l'identification de trois périodes-clés dans l'histoire du commerce international : 1) 1850-1914, période de l'industrialisation ; 2) 1945-1970, période du boom économique ; 3) 1985-2020, période de l'éclosion de nouvelles techniques de vente, du développement du marketing et de l'émergence du commerce électronique.

Le critère textuel nous a permis d'obtenir des données provenant de typologies textuelles différentes – textes institutionnels, textes scientifiques ou académiques, textes relevant de la presse spécialisée et textes issus de la documentation d'entreprise –, alors que le critère thématique a garanti la représentativité d'un large éventail de domaines, regroupés en quatre catégories (macroéconomie et économie internationale, produits et services) et quatre types d'activités (management, marketing, logistique, commerce électronique et droit).

Notre corpus se compose, au total, de 586 textes et 264.585 tokens¹. Afin de garantir une continuité dans le temps des textes choisis, ainsi que plusieurs autres activités les concernant, dont des recherches à partir de paramètres divers, le corpus a été stocké dans la plateforme Alfresco (<https://www.alfresco.com/it/piattaforma>) : il s'agit d'une plateforme open-source, adaptée aux besoins de notre corpus et de sa consultation de la part de notre équipe de techniciens.

Les critères qui ont présidé à l'établissement du corpus se retrouvent dans la plateforme comme autant d'éléments de filtrage : il est possible de consulter des textes relevant de la première, de la deuxième ou de la troisième période, sur l'une des activités identifiées, selon le type de texte ; ou, encore, nous avons décliné dans la plateforme les différents types de produits et de services, en les associant à tel ou tel autre texte ; en outre, les textes sont classés par mots-clés, qui figurent également dans le menu de gauche et qui peuvent être activés en phase de filtrage. Autre aspect activable pour le filtrage : le pays francophone dont il est question dans le texte.

Cette plateforme est à l'usage exclusif des membres de l'équipe DIACOM-fr et ne concerne que des recherches portant sur les textes du corpus : il ne s'agit donc pas de recherches concernant des contextes à partir des termes donnés, qui affectent plutôt la section qui suit (Section 3).

¹ Nous tenons à préciser que la totalité du corpus n'est pas accessible dans sa version publique (voir Section 3) car nous n'avons pas obtenu, pour la totalité des textes, l'autorisation de la part des auteurs. Soulignons, au passage, que tous les textes ont été soumis à un traitement prévoyant, d'abord, le passage du format .PDF au format .TXT et, ensuite, le nettoyage du bruit s'étant glissé dans les documents à l'issue de la conversion.

The screenshot shows a web interface for 'DIACOM FR Documents'. On the left, there are several filter sections: 'Active', 'Available', 'Texte intégral public' (with 'No (455)' and 'Yes (13)' options), 'Mots-clés', 'Type de document', 'Macroéconomie et économie internationale', 'Pays - Publication ou rédaction', 'Produits', 'Services', 'Entreprise', and 'Intervalles chronologiques'. The main area displays a table of documents with columns: Titre, Auteur, Texte intégral..., Date de publi..., Date de tom..., Mots-clés, Métriques..., Produits, Services, and a final column with letters like 'M'. The table contains several rows of document information.

Titre	Auteur	Texte intégral...	Date de publi...	Date de tom...	Mots-clés	Métriques...	Produits	Services	
Revenus		No	1997	2018	consommati...	Social issues			
Commerce e...	OMC	No	1999	2018	OMC, comm...	Politique co...			
Rapport sur l...	Organisation ...	No	2016	2018	OMC, PME, ...	Politique co...			
Orienter les e...	Chambre de...	No	2015	2018	Canada, PM...	Politique co...			
Peut-on iden...	Jean-Bernar...	No	2014	2018	DSGE, politi...	Politique co...			
Prévoir la co...	Jean Cristop...	No	1990	2018	conjoncture, ...				M
Primes de ris...	Patrick Arlus	No	1990	2018	prime de risq...	Politique co...			
Promotion d...	Equipe Cana...	No	1998	2018	Canada, co...	Politique co...			
Qualification ...	Patrick Seve...	No	1990	2018	main-d'œuvr...				M
Questions su...	Gilbert Cotte...	No	1990	2018	industrie ma...				M
Rapport d'ac...	Direction du ...	No	2014	2018	France, budg...				M

Figure 1
Corpus DIACOM-fr stocké sur la plateforme Alfresco.

3. Interrogation du corpus

La création du corpus a plusieurs finalités : la première est de rendre ce corpus disponible pour un ensemble relativement vaste d'utilisateurs (chercheurs, terminologues, terminographes, experts du domaine, enseignants, étudiants etc.) pour qu'ils puissent mener des recherches ad hoc concernant les données linguistiques. Pour que cet objectif se réalise, nous nous sommes servis d'un logiciel, lui aussi open source : il s'agit de KonText (voir Machálek 2020) que nous avons adapté à nos besoins. KonText est un moteur de recherche qui se caractérise par « a powerful query language, analytical functions, user-friendly interface, performance scalable to billion tokens corpora and tens or more concurrent users » (Machálek 2020, p. 7003).

Le masque de recherche se présente de manière très simple et donne la possibilité de rechercher les concordances à partir d'un terme donné, comme par exemple *commerce* :

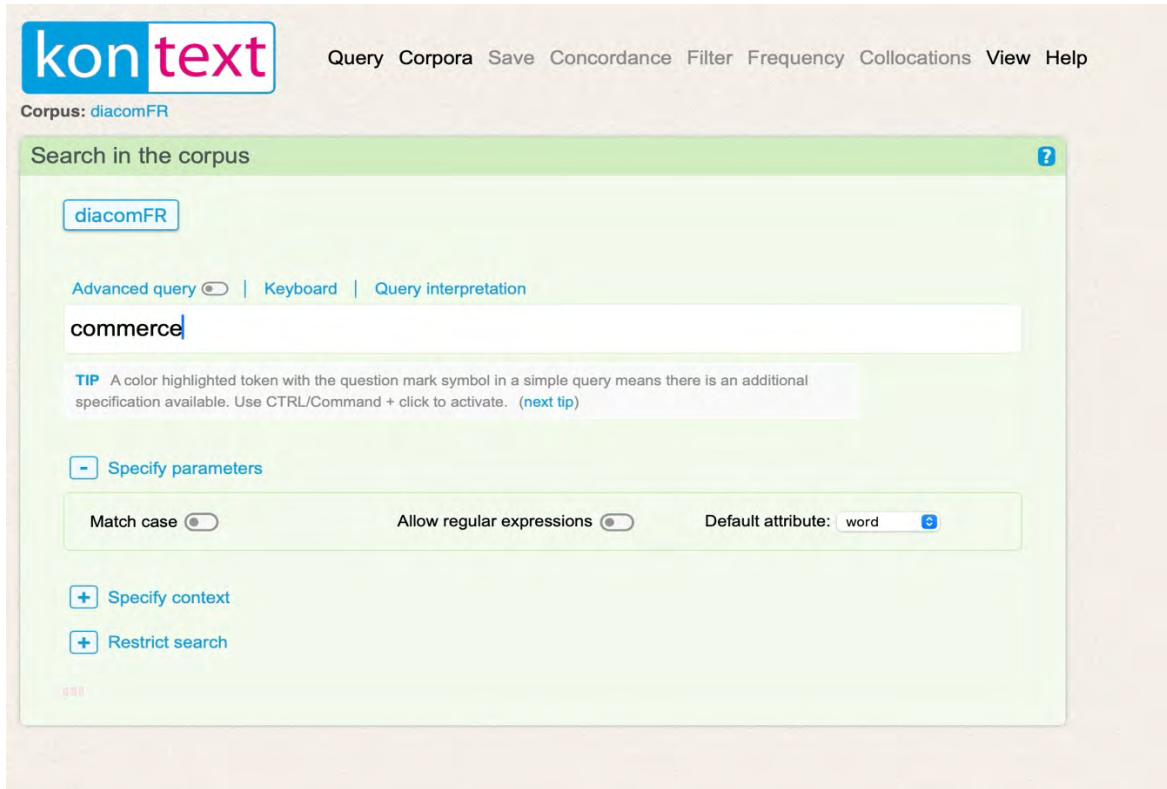


Figure 2
Masque de saisie à partir de Kontext pour DIACOM-fr.

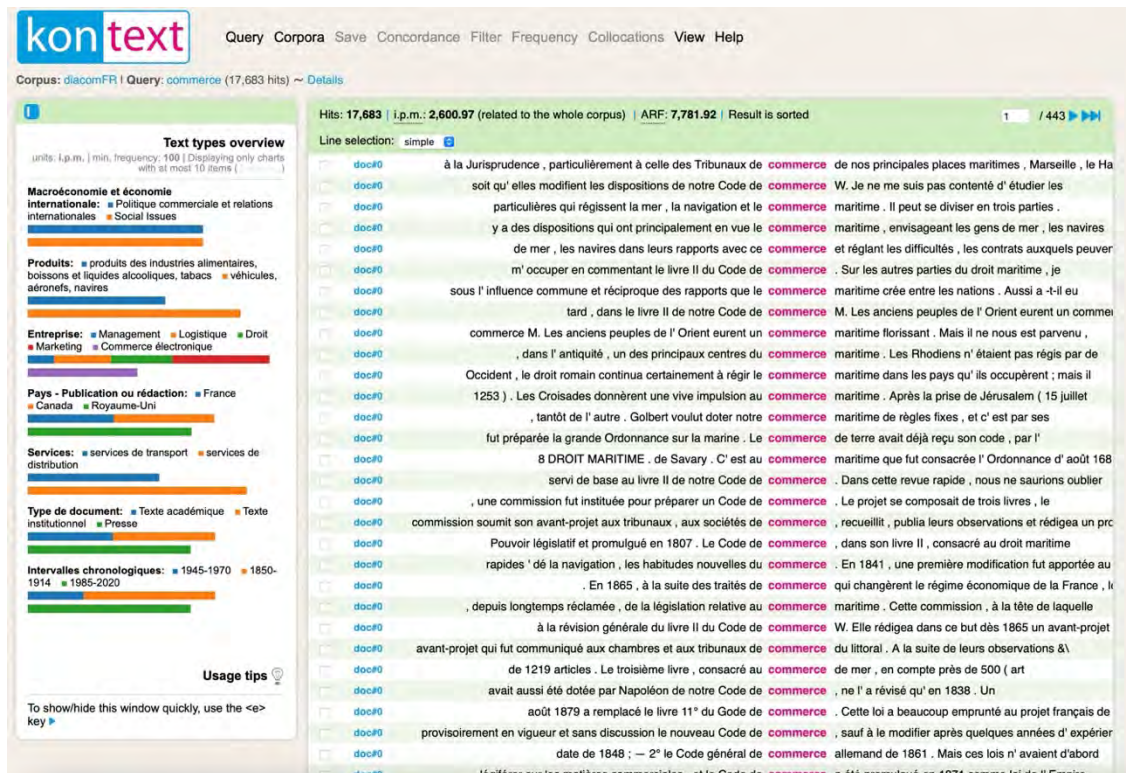


Figure 3
Concordances de *commerce* dans Kontext pour DIACOM-fr.

Pour ce qui est des concordances, en plus des contextes du terme recherché, KonText offre un ensemble de données statistiques à partir des paramètres adoptés pour la constitution du corpus. En particulier, le masque de gauche « Text types overview » fait état des proportions des sous-domaines principaux, des types de produits et d'entreprise, des pays où les textes ont été publiés, des types de services, des types de documents et des intervalles chronologiques.

Nous précisons aussi que, comme dans la plupart des logiciels de ce type, la longueur des contextes à gauche et à droite du terme recherché est paramétrable.

En outre, les contextes peuvent être modifiés à l'aide des filtres contenus dans « Restrict Search » : ici, il sera possible de restreindre les contextes selon le type de texte, la période historique, le type de produit ou de service etc. – il s'agit par ailleurs des mêmes critères enregistrés dans la plateforme Alfresco (voir Section 2) – comme montré dans la figure qui suit :

The screenshot shows the KonText interface for the 'diacomFR' corpus. The 'Restrict search' section is expanded, showing several filter categories:

- Titre**: Macroéconomie et économie internationale
- Auteur**: Produits
- Date de publication**: Source
- Entreprise**: Pays - Publication ou rédaction
- Services**: Services
- Type de document**:

<input type="checkbox"/>	161,908
<input type="checkbox"/> Presse	273,146
<input type="checkbox"/> Presse/Texte académique	20,083
<input type="checkbox"/> Texte académique	1,600,148
<input type="checkbox"/> Texte institutionnel	5,126,948
<input type="checkbox"/> Texte institutionnel/Texte académique	343,455
- Intervalles chronologiques**:

<input type="checkbox"/> ===NONE===	2,780
<input type="checkbox"/> 1850-1914	4,353,483
<input type="checkbox"/> 1945-1970	260,157
<input type="checkbox"/> 1985-2020	2,182,192

Figure 4
Paramètres concernant la « Restrict Search » dans KonText pour DIACOM-fr.

4. Extraction des données

Même si la base de données se veut d’abord synchronique et qu’elle se focalise davantage sur les liens syntagmatiques et paradigmatiques qui caractérisent la période actuelle (voir Section 4.2), puisque les liens diachroniques font aussi partie des liens pris en compte, nous avons effectué l’extraction terminologique distinctement pour les trois périodes, en obtenant ainsi trois nomenclatures différentes.

À partir du corpus constitué, nous avons procédé à l’extraction des termes simples et complexes. Pour l’extraction des termes, nous nous sommes appuyés principalement sur le logiciel TermoStat, conçu et élaboré à l’Université de Montréal par Patrick Drouin (voir Drouin 2010).

Ce logiciel est particulièrement performant et s’adapte à la typologie de corpus et de données attendues. La performance de TermoStat a été mesurée par un test effectué sur plusieurs extracteurs et logiciels actuellement disponibles : AntConc, Sketch Engine, TermoStat et TermSuite.

Par rapport aux autres logiciel, TermoStat présente les avantages suivants (voir également Calvi, Drouin, Frassi 2023) :

- Il est relativement fiable pour ce qui est de l’extraction des unités multilexémiques, notamment des collocations. En effet, les termes complexes sont surtout des syntagmes nominaux. Néanmoins, certaines associations, comme les collocations à collocatif verbal, qui se caractérisent souvent par un verbe en première position (ex. : *dresser un bilan*), sont difficilement identifiables par la plupart des extracteurs/logiciels ; or, TermoStat permet l’identification de ces types de termes complexes grâce à la fonction *bigrammes*. Cette fonction n’existe pas dans les autres logiciels : dans AntConc (avec la fonction *Collocate*) et Sketch Engine (avec la fonction *Word sketches*), par exemple, on peut retrouver les collocations à collocatif verbal uniquement en insérant la base dans le masque de saisie ; ces logiciels ne fournissent pas une liste de collocations de ce type à partir d’une extraction automatique. Cet aspect s’avère utile pour nous car nous ne considérons pas les collocations comme des termes à proprement parler, mais comme une propriété de combinatoire des termes (voir à ce propos la Section 4.1).
- L’intérêt de TermoStat réside aussi dans le corpus de référence dont il se sert et de la manière d’identifier automatiquement le statut plus ‘terminologique’, si l’on peut dire, d’un corpus ou d’une liste de mots simples ou complexes. Ce logiciel reçoit un corpus en entrée et, à l’issue d’une comparaison avec un corpus de référence non spécialisé, il propose de manière automatique une liste de candidats-termes simples et/ou complexes. S’appuyant sur l’hypothèse que les termes d’un domaine donné sont bien plus fréquents dans un corpus spécialisé du même domaine que

dans le corpus de langue générale, la mise en opposition de deux corpus joue un rôle d'indéniable importance pour la réduction du bruit.

Ces deux avantages ont été pour nous fondamentaux dans le repérage des termes qui peuplent notre base de données, car ils fournissent des listes sur lesquelles le travail manuel de filtrage – qui est indispensable – a été simplifié. Nous avons intégré nos nomenclatures par les termes simples et complexes figurant dans l'un des ouvrages terminographiques de référence pour les trois périodes. Ces ouvrages résultent d'un dépouillement concernant tous les ouvrages terminographiques du domaine sur les trois périodes, qui nous a donné 5 ouvrages pour la première tranche, 22 pour la deuxième et 79 pour la troisième. Nous avons retenu 2 ouvrages pour chaque période, choisis en raison de la richesse et de l'exhaustivité de leurs nomenclatures : il s'agit de Sacré (1884) et Guyot et Rafalovich (1901) pour la période allant de 1850 à 1914 ; de Kettridge (1949) et Centre international du droit des affaires (1973) pour la période allant de 1945 à 1970 ; de Massabié-François et Poulain (2002) et Morelle (2003) pour la période allant de 1985 à 2020. À l'issue de cette extraction, nous avons retenu 5030 termes (simples et complexes).

5. Les propriétés des termes

Une fois les données identifiées, il est nécessaire d'établir le modèle nous permettant de les représenter dans une base de données.

Les bases de données terminologiques existent depuis longtemps ; dans le cadre de notre projet, nous avons choisi de nous appuyer sur un modèle de représentation des données terminologiques assez intuitif et, surtout, respectueux des propriétés paradigmatiques et syntagmatiques des unités terminologiques, ainsi que des propriétés syntactico-sémantiques de chaque entité terminologique. Cela nous a poussé à une réflexion concernant 1) la définition des entités terminologiques et leurs propriétés syntactico-sémantiques ; 2) les propriétés paradigmatiques et syntagmatiques qui les caractérisent, ainsi que d'autres propriétés que le terminographe souhaite valoriser ; 3) le choix d'un modèle dans lequel ces aspects soient représentés de manière intuitive, en vue d'une consultation rapide de la part de tout utilisateur ainsi que dans des objectifs didactiques.

Nous croyons que le modèle plus traditionnel de la base de données – celui qui emprunte à la lexicographie et à la terminographie classiques le caractère linéaire – ne s'adapte pas à nos objectifs ; ainsi, nous avons préféré de nous tourner du côté du modèle des réseaux lexicaux (voir à ce propos Polguère 2014). Ce modèle représente les entités lexicales en respectant leurs propriétés : il se configure comme un graphe lexical dans lequel chaque nœud correspond à une entité lexicale et des liens se tissent entre nœuds, représentant

autant de propriétés paradigmatiques ou syntagmatiques des diverses entités terminologiques.

Ce modèle a été déjà appliqué dans la langue générale : il suffit de penser à WordNet ou FrameNet, pour la langue anglaise, ou, encore, au *Réseau lexical du français* (RL-fr, <https://spiderlex.atilf.fr/fr>) pour ce qui est de la langue française.

Cette dernière expérience lexicographique est notre modèle d'inspiration pour la base de données terminologiques que nous avons constituée.

En amont de la création de cette base il est nécessaire d'identifier :

- 1) Les entités terminologiques ;
- 2) Les types de liens que ces entités nouent entre elles ;
- 3) Les autres propriétés à représenter.

5.1. Les entités terminologiques

Du point de vue linguistique, les unités de base de la terminologie s'appellent *termes* : il s'agit de formes graphiques rattachées à un signifié et appartenant clairement à un domaine de spécialité.

La tradition terminologique et terminographique distingue deux types de termes : les termes simples et les termes complexes (voir par exemple L'Homme 2020 et Cabré 1992).

Nous avons voulu aller plus loin dans l'identification des termes complexes et, à partir de deux études pilotes sur deux sous-corpus différents (les résultats détaillés de ces deux études se trouvent dans Frassi, Calvi, Humbley 2020 ; Frassi 2020, 2023), nous avons été en mesure de démontrer qu'un grand nombre de termes complexes partage les propriétés syntactico-sémantiques qui ont été traditionnellement rattachées aux locutions, notamment par Gross (1996) et Mel'čuk (2013). Plus particulièrement, pour notre base de données nous avons adopté le classement de Mel'čuk (2013), identifiant des cas de locutions fortes, semi-compositionnelles et faibles.

Les locutions ne sont pas, pour autant, les seuls types d'unités multilexémiques apparaissant en terminologie : notre étude pilote a révélé une quantité non négligeable de collocations. Puisque les collocations en terminologie possèdent essentiellement les mêmes propriétés que les collocations dans la langue générale, les mêmes types de liens syntagmatiques se présentent en domaine de spécialité : par exemple, si nous reprenons le répertoire des liens proposés dans le cadre de la Lexicologie Explicative et Combinatoire (voir Mel'čuk, Clas, Polguère 1995 ; Mel'čuk, Polguère 2022 ; voir également la Section 4.2), il est possible d'identifier le sens de 'bon', dans les collocations *rapport fiduciaire* ou *client fidèle*, ou, pour ce qui est des collocations à collocatif verbal, des verbes de réalisation comme dans *réaliser un investissement*. Toutefois, nous ne considérons pas la collocation comme

une entité terminologique ; elle exprime une propriété de combinatoire de la base (terme simple ou terme complexe).

Ainsi, notre base de données va considérer les entités terminologiques correspondant à des termes simples (à savoir des mots graphiques, suites de lettres séparées par des espaces blancs, comme par exemple *commerce*) et à des termes complexes de type *locutions*.

5.2. Les liens entre termes

Les graphes lexicaux vont tenir grâce à des liens paradigmatiques et syntagmatiques. Pour l'identification de ces liens, nous nous sommes appuyés sur le répertoire dressé dans le cadre de la Lexicologie Explicative et Combinatoire (Mel'čuk, Clas, Polguère 1995 ; Polguère, Mel'čuk 2022) : ces liens, très généraux et très récurrents dans la langue générale, se trouvent également dans les langues de spécialité. Ainsi, par exemple, le lien de synonymie (*commerce/transaction*) ou le nom typique de l'un des actants (*commerce/vendeur/acheteur/marchandise/prix...*).

5.3. Autres propriétés des termes

D'autres aspects vont être pris en compte en tant que propriétés des termes. Il s'agit notamment de la diachronie et des aspects didactiques.

5.3.1. La diachronie

Les liens diachroniques permettent de relier, sur l'axe du temps, des unités terminologiques appartenant à des périodes différentes. Ces liens se créent au niveau de la forme et/ou du sens et peuvent être mis au nombre des liens de type paradigmatique.

Pour marquer l'appartenance de tel ou tel autre terme à l'une des trois périodes envisagées, nous avons adopté une notation impliquant l'emploi des trois lettres de l'alphabet grec α (pour la période 1985-2020), β (pour la période 1945-1970), γ (pour la période 1850-1914). Chaque terme pouvant être potentiellement polysémique, il est aussi accompagné d'un chiffre romain indiquant l'acception. Ainsi, par exemple, *commerce α 1* réfère à la première acception du terme *commerce* dans la période 1985-2020.

Pour revenir à la question des liens diachroniques entre termes, comme nous l'avons déjà souligné dans Frassi (2021), nous avons identifié, avec l'aide de plusieurs études concernant la terminologie diachronique, et notamment Møller (1998), Condamines, Rebeyrolles, Soubeille (2004), Picton (2014) et Dury (2018), les cas de variation les plus récurrents, et plus précisément :

- 1) Expansion (ex. *billet à ordrey1* → *billet à ordrea1* → *billet à ordre international α 1*) : un terme simple devient complexe ou un terme complexe ajoute un élément ultérieur ; cette variation implique généralement une

précision dans le sémantisme du terme à l'issue de cette expansion formelle.

- 2) Ellipse ou réduction (ex. *protection contre les rayonnements* → *radioprotection*²): un terme complexe réduit le nombre de ses composantes ; cette variation implique généralement une identité dans le sémantisme des deux termes, l'ellipse se configurant comme une opération de simplification s'appliquant uniquement à la forme.
- 3) Resémantisation (ex. *transbordement*₁ → *transbordement*₁): un terme demeure tel quel du point de vue de la forme mais subit une variation au niveau du signifié.

5.3.2. Les 'propriétés' didactiques³

Comme nous l'avons rappelé plus haut, en plus des propriétés traditionnelles, nous poursuivons également l'exploitation des données terminologiques dans un contexte didactique. C'est sur ce dernier aspect que nous focalisons maintenant notre attention.

Chaque terme vedette faisant partie de notre base de données a été associé à un niveau d'apprentissage, et ce dans l'objectif d'aider enseignants et apprenants dans l'exploitation de la base de données elle-même à des fins didactiques. Un champ à part entière, appelé *niveau d'apprentissage*, intègre de ce fait la fiche terminologique du DIACOM-fr. À chaque terme est associé l'un des trois niveaux d'apprentissage identifiés, *élémentaire*, *indépendant* et *expérimenté*, qui correspondent aux trois niveaux généraux de la compétence linguistique établis par le CECR.

Pour attribuer à chaque terme un niveau de compétence, nous avons adopté un critère principalement qualitatif (voir, à ce propos, Frassi et Bonadonna 2022 pour le classement des termes de type *locutions* ; Frassi et Bonadonna 2023 pour le classement des acceptions de la polysémie). En particulier, nous avons associé à l'échelon le plus bas (A2) des termes simples dont les sens pouvaient être considérés comme *de base* (à savoir typiquement associés au domaine du commerce, comme par exemple *commercial* = 'relatif au commerce'). Dans le cas de termes polysémiques, il s'agissait toujours de la première acception, qui était également, du point de vue quantitatif et dans la grande majorité des cas, la plus fréquente. L'échelon intermédiaire, correspondant aux niveaux B1/B2, inclut les termes simples dont l'acception n'est pas considérée comme une acception de base, ainsi que les locutions qui, du point de vue sémantique, sont semi-compositionnelles. Pour ce qui est, finalement, du niveau C1, y sont associés les termes simples dont l'acception,

² Nous empruntons cet exemple à Goffin (1989, p. 105).

³ Plus en général, sur l'exploitation des corpus en didactique du FOS – notamment dans les domaines du commerce et du marketing digital – voir Bonadonna 2023.

puisqu'elle est métaphorique ou métonymique, demande un effort particulier de la part de l'apprenant ; quant aux termes complexes de type locution, il s'agit d'unités multilexémiques non compositionnelles (ex. : *marge arrière*).

Il va de soi que, selon les critères de classement adoptés pour la propriété didactique, la plupart des termes sont des termes de base ou des acceptions de base non métaphoriques ni métonymiques, ou des locutions faibles (compositionnelles du point de vue sémantique).

6. La représentation dans la base des données

Dans le présent paragraphe nous faisons état de la méthode choisie pour la représentation des données terminologiques et de leurs propriétés. Ces choix se trouvent déjà dans Frassi et Rospocher (2022) : nous nous bornons ici à rappeler les aspects de base.

Notre approche s'est basée sur l'adoption des technologies du web sémantique via le modèle RDF pour la représentation des données, de leurs propriétés et de leurs liens. Plus spécifiquement, pour la modélisation des entités terminologiques nous nous sommes appuyés sur le modèle Lemon (McCrae *et al.* 2017). Ce modèle a été exploité pour la représentation des entités terminologiques (termes simples ou termes complexes de type *locution*), chacune étant associée à un signifié spécifique et à une partie du discours. Pour la modélisation des liens paradigmatiques et syntagmatiques, ainsi que pour les liens de type diachronique, nous avons adopté le modèle Lexfom (Lexical Functions Ontology Model, Fonseca *et al.* 2016). Celui-ci a été conçu pour la modélisation des fonctions lexicales selon la Théorie Sens-Texte et la Lexicologie Explicative et Combinatoire.

Ces modèles nous permettent ainsi de représenter des entités terminologiques et leurs propriétés sous forme de web sémantique prévoyant une interrogation à partir des entités lexicales qui constituent notre nomenclature.

Les entités terminologiques simples et complexes (celles-ci uniquement de type *locution*) s'affichent à l'écran entourées de leurs propriétés diverses (liens paradigmatiques et syntagmatiques, propriétés diachroniques et propriétés didactiques), comme dans les deux images qui suivent, représentant un terme simple (*transaction*) et un terme complexe de type *locution* (*transaction courante*) :

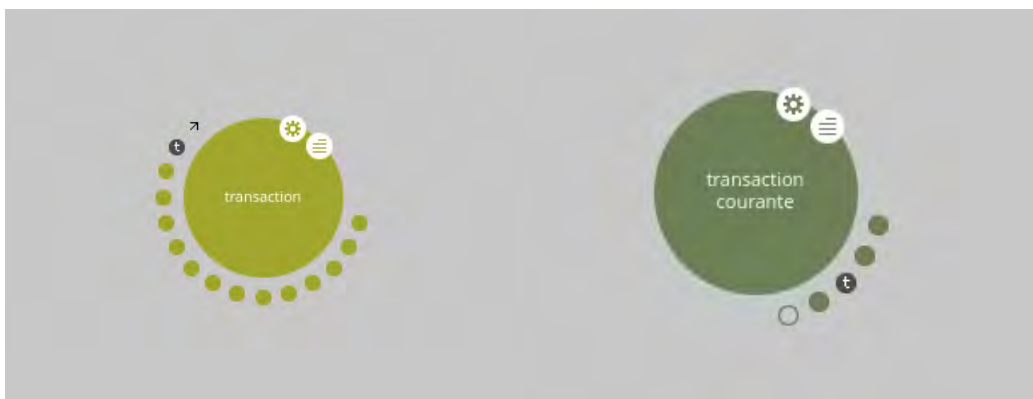


Figure 5
Exemple de représentation des entités terminologiques dans DIACOM-fr.

Les propriétés des termes s'affichent sous forme de petits ronds entourant l'entité terminologique elle-même. Ainsi, le nombre de petits ronds dépend du nombre de propriétés valorisées pour chaque entité terminologique : en l'occurrence, *transaction* développe un réseau de liens plus vaste que *transaction courante*. En cliquant sur chacune des propriétés, il est possible d'afficher le type de lien et l'entité vers laquelle ce lien pointe.

Ainsi, par exemple, pour ce qui est de *transaction courante*, en cliquant la propriété diachronique, on aura le réseau suivant, qui atteste la présence de ce même terme dans les deux autres périodes envisagées dans notre diachronie, avec une relation d'identité de sens et de forme :

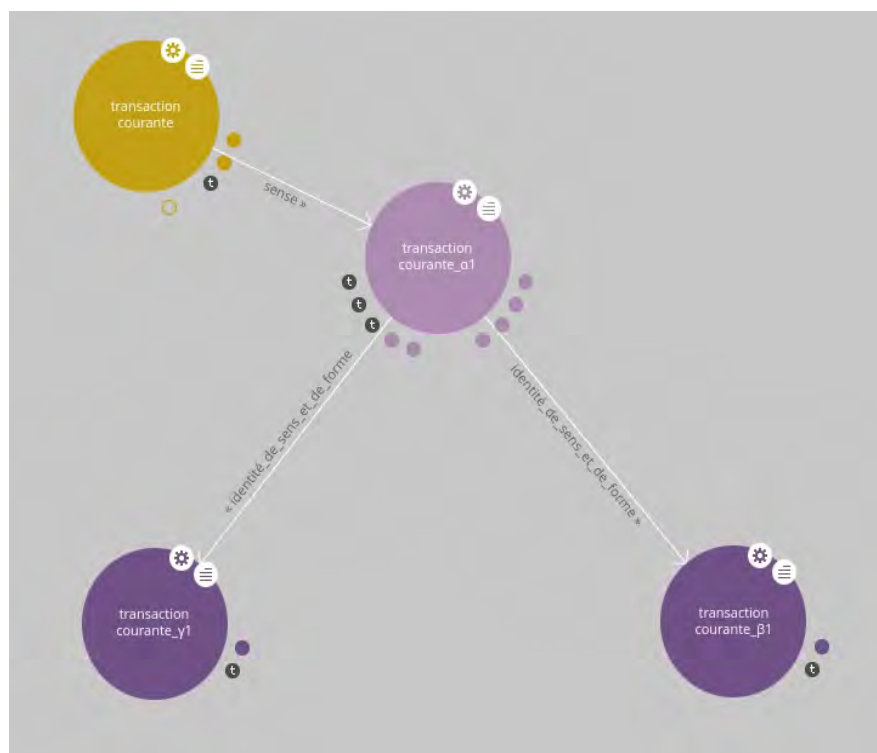


Figure 6
Les propriétés diachroniques d'un terme dans DIACOM-fr.

Quant aux liens syntagmatiques et paradigmatiques, nous proposons l'exemple du réseau qui se tisse à partir de *transaction* :

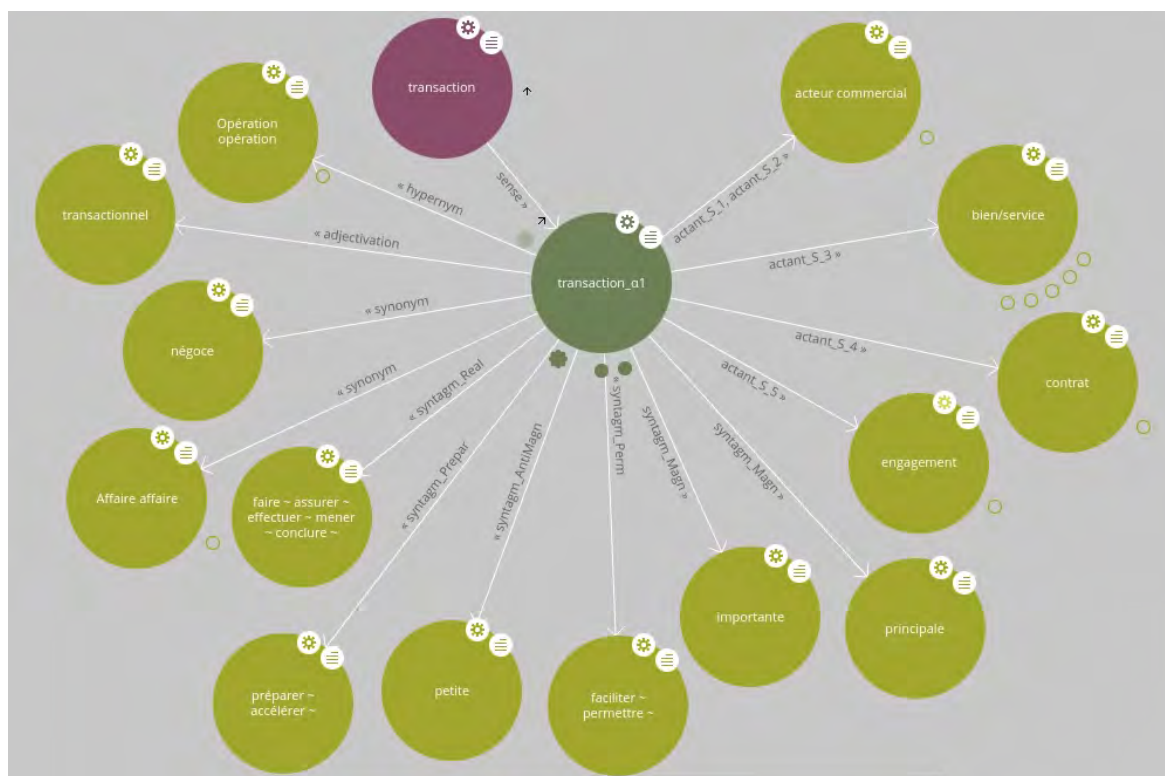


Figure 7

Affichage en réseau des liens paradigmatiques et syntagmatiques (exemple de *transaction*), DIACOM-fr.

Les relations paradigmatiques sont représentées par les liens d'hyperonymie (« hypernym », *opération commerciale*), d'adjectivisation (*transactionnel*) et de noms d'actants (*acteur commercial*, *bien/service*, *contrat* et *engagement*).

Les relations syntagmatiques – précédées de la chaîne de caractères « syntagm_ » - concernent, pour le cas de *transaction*, des verbes de réalisation (*faire*, *assurer*, *effectuer*, *mener*, *conclure une transaction*), des verbes de préparation (*préparer*, *accélérer une transaction*), des verbes exprimant le sens 'ne pas faire cesser, laisser continuer' (*faciliter*, *permettre une transaction*), les modificateurs d'intensité (*transaction principale*, *transaction importante*) et les modificateurs exprimant le sens opposé à l'idée d'intensité (*petite transaction*).

Les propriétés didactiques ne sont pas encore valorisées dans la base : il s'agit de la dernière étape de notre activité terminographique en phase d'achèvement.

7. Conclusion

Nous avons fait le tour d'un projet terminologique ayant plusieurs débouchés et, par conséquent, plusieurs points d'intérêt.

Premièrement, du point de vue de la linguistique de corpus, notre projet permet la consultation d'un corpus récent à partir d'un logiciel – KonText – adapté à cette fin et permettant le paramétrage concernant des critères différents affectant, par exemple, la diachronie, le domaine, le type de texte, le type de produit ou service, etc.

L'aspect le plus novateur est représenté par la méthodologie adoptée pour l'affichage du réseau terminologique lui-même, incluant une réflexion de type linguistique et une réflexion touchant à l'informatique. Au niveau linguistique nous avons identifié non seulement les types principaux d'entités terminologiques – termes simples et termes complexes de type *locution* – mais nous avons également motivé notre choix de ne pas faire figurer, en tant que lemmes dans la base elle-même, les collocations. En outre, nous avons isolé un ensemble de propriétés des termes, dont certaines ne sont pas communément représentées dans les bases de données terminologiques : en plus des liens paradigmatiques et syntagmatiques, en effet, nous nous sommes occupés d'isoler les types de liens intervenant au niveau de la diachronie et de classer les termes selon le niveau de compétence linguistique. Ces derniers aspects sont particulièrement novateurs dans le domaine terminologique. Quant aux premiers, nous nous sommes appuyés sur les types de liens identifiés par la Lexicologie Explicative et Combinatoire, ce qui constitue une tentative certes non-pionnière, mais assez peu répandue.

La méthodologie adoptée pour adapter ces données linguistiques et ces propriétés linguistiques à une base de données informatisée demeure aussi intéressante : nous avons, en effet, réfléchi non seulement à la modélisation des types de propriétés de chaque terme, à travers l'exploitation de Lexfom, mais nous proposons une modélisation, en général, de toutes les données linguistiques à travers de triplets (RDF). Une telle modélisation répond à l'idée général du web sémantique et permet de représenter les données non pas à travers un affichage linéaire et traditionnel, mais à travers un graphe lexical qui permet des avantages de consultation : davantage intuitif, il permet sans doute un meilleur accès aux propriétés des termes et aux liens entre termes.

La base de données DIACOM-fr s'adresse à des types divers d'utilisateurs, dont des linguistes, des terminologues et des terminographes, des enseignants et des apprenants : nous souhaitons que la conformation de notre réseau terminologique puisse leur rendre service.

Bionote : Paolo Frassi est Professeur de Linguistique française à l'université de Vérone. Ses intérêts scientifiques concernent la lexicologie et la lexicographie, ainsi que la terminologie,

la terminographie, l'enseignement du FLE. Il a consacré de nombreuses études à la définition de l'adjectif ainsi qu'au statut des phrasèmes entre langue générale et terminologie. Au sein du Département des Langues et Littératures étrangères de l'Université de Vérone, il a dirigé le projet d'envergure 2018-2022 sur « Les humanités numériques appliquées aux langues et littératures étrangères ». Il a participé à plusieurs projets, dont DIACOM-fr (Université de Vérone), RL-fr (ATILF-CNRS, Nancy), FLATIF (PRIN) et collabore avec le Laboratoire ATILF-CNRS de Nancy ainsi qu'avec l'UQAM et l'Université de Montréal.

Adresse de l'auteur : paolo.frassi@univr.it

Références bibliographiques

- Alfresco <https://www.alfresco.com/it/piattaforma> (25/10/2023).
- Bonadonna, M.F. 2023, *Didactique du lexique et corpus numériques pour le Français L2. Des applications pour le commerce et le marketing digital*, Peter Lang, Bruxelles.
- Cabré M.T. 1992, *La terminologia. La teoria, els mètodes, les aplicacions*, Empúries, Barcelona.
- Calvi S., Drouin P. et Frassi P. 2023, *Acquisition (semi-)automatique des collocations terminologiques*, in Frassi P. (éd.), *Phraséologie et terminologie*, De Gruyter, Berlin, Boston, pp. 239-258.
- Centre International du Droit des Affaires 1973, *Lexique pratique commercial*, Régif, Paris.
- Condamines A., Rebeyrolle J. et Soubeille A. 2004, *Variation de la terminologie dans le temps : une méthode linguistique pour mesurer l'évolution de la connaissance en corpus*, in Williams G. et Vessier S. (éds.), *Proceedings of the 11th EURALEX international Congress, Euralex 2004*, Lorient, 6-10 juillet 2004, Université de Bretagne-Sud, Faculté des Lettres et des Sciences humaines, Lorient, pp. 547-557.
- Drouin P. 2010, *TermoStat. Présentation du logiciel* http://termostat.ling.umontreal.ca/doc_termostat/doc_termostat.html (25/10/2023).
- Dury P. 2018, *La dimension diachronique en anglais de spécialité : une approche terminologique*, Note de synthèse pour l'habilitation à diriger des recherches, Université Paris 7.
- Fonseca A., Sadat F. et Lareau F. 2016, *Lexfom: a lexical functions ontology model*, in *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V)*, pp. 145-155.
- Frassi P., Calvi S. et Humbley J. 2020, *Fouille de textes et repérage d'unités phraséologiques*, in *Terminologie & ontologie : théorie et applications*, Presses Universitaires Savoie Mont Blanc, Chambéry, pp. 321-338.
- Frassi P. 2020, *La force des locutions faibles en domaine de spécialité*, in Zanola M.T. et Conceição M.C. (éds.), *Terminologia e mediação linguística, métodos, práticas e atividades*, UAlg editora, Faro, pp. 43-60.
- Frassi P. 2021, *DIACOM-fr, une base de données terminologiques de type diachronique*, in « Cahiers de Lexicologie » 118 [1], pp. 23-50.
- Frassi P. 2023, *Les unités multilexémiques non libres entre langue générale et langue de spécialité*, in Henrot Sostero G. (éd.), *Alle radici della fraseologia europea*, Peter Lang, Lausanne, pp. 281-304.
- Frassi P. et Bonadonna M.F. 2022, *Les termes complexes de type locution dans l'enseignement du français L2 en langue de spécialité : le cas du domaine du commerce international*, in « LIDIL » 65, pp. 1-19.
- Frassi P. et Bonadonna M.F. 2023, *Termes, polysémie et niveaux d'apprentissage en FLE*, in « Recherches et applications. Le Français dans le Monde » 73, pp. 57-74.
- Frassi P. e Rospocher M. 2022, *Ontologie, locuzioni deboli e legami fra entità terminologiche nell'ambito del progetto DIACOM-fr*, in Chiocchetti E. e Ralli N. (éds.), *Risorse e strumenti per l'elaborazione e la diffusione della terminologia in Italia*, Eurac Research, Bolzano, pp. 138-151.
- Goffin R. 1989, *La terminologie des sciences et des techniques nucléaires. Un cas de diachronie récente*, in Centre de Terminologie de Bruxelles (éd.), *Terminologie diachronique. Actes du colloque organisé à Bruxelles les 25 et 26 mars 1988*, Conseil international de la langue française, Paris, pp. 94-107.

- Gross G. 1996, *Les expressions figées en français : noms composés et autres locutions*, Ophrys, Paris.
- Guyot Y. et Rafalovich A.G. 1901, *Dictionnaire du commerce, de l'industrie et de la banque : Assurance, banque et bourse, comptabilité, droit commercial, terrestre et maritime, économie politique appliquée*, Guillaumin, Paris.
- Kettridge J.O. 1949, *French-English/English-French Dictionary of Commercial and Financial Terms. Phrases and Practice*, Routledge and Kegan Paul Limited, London.
- KonText <http://kontext.korpus.cz/> (25/10/2023).
- L'Homme M.-C. 2020, *La terminologie : principes et techniques*, Les Presses de l'Université de Montréal, Montréal.
- Machálek T. 2020, *KonText: Advanced and Flexible Corpus Query Interface*, in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, pp. 7003-7008.
- Massabié-François M. et Poulain E. 2022, *Lexique du commerce international. Les 2 600 mots actuels et pratiques de l'import-export*, Bréal, Rosny-sous-Bois.
- Mel'čuk I. 2013, *Tout ce que nous voulions savoir sur les phrasèmes mais...*, in « Cahiers de lexicologie » 102 [1], pp. 129-149.
- Mel'čuk I., Clas A. et Polguère A. 1995, *Introduction à la lexicologie explicative et combinatoire*, Duculot, Paris.
- Mel'čuk I. et Polguère A. 2021, *Les fonctions lexicales dernier cri*, in Marengo S., *La Théorie Sens-Texte et ses applications. Lexicologie, lexicographie, terminologie, didactique des langues*, L'Harmattan, Paris, pp. 75-155.
- McCrae J.P., Bosque-Gil J., Gracia J., Buitelaar P. and Cimiano P. 2017, *The Ontolex-Lemon Model: Development and Applications*, in *Proceedings of eLex 2017 Conference, 19–21 September 2017, Leiden, Netherlands*, Lexical Computing CZ s.r.o, Leiden, pp. 587-597.
- Møller B. 1998, *À la recherche d'une terminochronie*, in « Meta » 43 [3], pp. 426-438.
- Morelle F. 2003, *Le Dico du commerce international* <https://www.glossaire-international.com/> (25/10/2023).
- Picton A. 2014, *The dynamics of terminology in short-term diachrony: a proposal for a corpus-based methodology to observe knowledge evolution*, in Temmerman R. and Van Campenhout M. (éds.), *The dynamics of culture-bound terminology in monolingual and multilingual communication, Terminology and Lexicography Research and Practice*, John Benjamins, Amsterdam/ Philadelphia, pp. 159-182.
- Polguère A. 2014, *From Writing Dictionaries to Weaving Lexical Networks*, in « International Journal of Lexicography » 27 [4], pp. 396-418.
- Sacré A. 1884, *Dictionnaire de commerce et de droit commercial*, Librairie de A. Marescq, Paris.
- Termostat Web 3.0+ <http://termostat.ling.umontreal.ca> (25/10/2023).

LES DÉNOMINATIONS DES PRODUITS TYPIQUES DE L'ÉMILIE-ROMAGNE

Diffusion multilingue d'une terminologie protégée

CHIARA PREITE
UNIVERSITÀ DI MODENA E REGGIO EMILIA

Abstract – This paper concentrates on Food and Wine terminology in the Emilia Romagna region, Italy. Because their form involves references to the geographical origins of the designated products, Food and Wine terms represent lexico-cultural denominations that contribute to shaping regional identity and promoting the territory. In a globalised world, local products and food traditions travel with their names, beyond local and national borders, to enter other communities, languages and cultures. In the light of this, we select from the *Sapori Locali* glossary the protected denominations of around fifty typical products from the Emilia-Romagna Region, and compare and contrast their forms with their French and English equivalents. Our goal is to assess the impact that their culturally specific territorial value may have on their international reception.

Keywords: terminology; protected denominations; typical products; multilingual equivalents.

1. Introduction

Dans ce nouveau millénaire, le secteur agroalimentaire jouit en Europe d'un grand intérêt en ce qu'il touche de différentes manières plusieurs domaines, dont Zanola (2015) en énumère de nombreux allant de la technologie pour la production et la transformation des produits, au marché qui les propose et diffuse ; de la qualité de ces mêmes produits, aux aspects juridiques et normatifs qui en protègent la production et la vente, jusqu'aux exigences médicales liées, par exemple, à la santé du consommateur, pour en arriver aux secteurs de la traduction, de la terminologie et de la promotion touristique.

En effet, des recherches ont foisonné également dans le champ de la linguistique et de la terminologie¹. La réalisation du *Glossaire multilingue*

¹ Sur le rapport entre gastronomie, langue, terminologie et culture, il est possible de consulter, entre autres, Cadeddu, Kauffer et Keromnes (2019) ; Chessa, De Giovanni et Zanola (2014) ; Marti, Planelles Iváñez et Sandakova (2018). Sans oublier des projets d'envergure tels AtLiTeG (<https://www.atliteg.org/>) ou AGROTERM (<https://centridiricerca.unicatt.it/optl-progetti-sostenibilita-ambientale-e-alimentare>).

*des produits typiques émiliens-romagnols Sapori locali*² se place dans ce cadre et conjugue les exigences de la transformation numérique engageant les études terminologiques et lexicographiques, avec la valorisation du patrimoine culturel local. Par ce glossaire, un aperçu des connaissances œnogastronomiques de la région italienne de l'Émilie-Romagne est mis à la disposition d'un public d'usagers passionnés ou de touristes, non seulement en italien, mais aussi en français, anglais, espagnol et roumain³.

La terminologie de l'œnogastronomie concerne aussi les procédés de culture, travail, transformation et conservation des produits, que l'on rencontre dans les notes descriptives et dans les contextes des fiches du glossaire (section 2). La présence de ces champs permet aux usagers, chacun dans sa langue, de découvrir des curiosités et des informations à propos des produits typiques émiliens-romagnols, vecteurs de la culture et du savoir-faire artisanal et industriel d'une région italienne, voire de l'*italianité* dont les cuisines régionales sont parmi les emblèmes. Toutefois, l'objectif de *Sapori locali* n'est pas d'adopter dans sa nomenclature ces termes techniques, mais seulement les dénominations des produits décrits.

Ces éléments lexicoculturels de nature alimentaire et territoriale sont partagés par la société (et la langue) qui les a conçus, traditionnalisés et diffusés. Ainsi ils concourent à la construction de l'identité sociale et participent d'un « sentiment collectif d'appartenance » (Fischler 2001, p. 68). En effet, puisque les produits pris en considération vantent des dénominations protégées (section 2), il est possible de les considérer de plein droit comme ayant un contenu territorial élevé⁴. Cela ressortit au premier regard, car la plupart de ces dénominations (section 3) montrent une localisation géographique spécifique, manifestée par la présence d'un toponyme, presque inséparable du produit. C'est le cas, pour porter quelques exemples parmi les plus célèbres, de l'« Aceto Balsamico di Modena », de la « Ciliegia di Vignola » ou du « Prosciutto di Parma ». L'association binaire *produit + ville*, que Galisson (1987, 1988, 1991) reconnaît parmi les vecteurs d'une charge culturelle partagée, devient l'indicateur d'une spécialité œnogastronomique italienne réputée par le biais duquel le produit typique promeut le terroir, et ce dernier à son tour valorise le produit.

² Projet conduit avec le soutien du Département d'Études Linguistiques et Culturelles de l'Université de Modène et Reggio d'Émilie et parrainé aussi par le Centre de recherche interdépartemental DHMoRe. La méthodologie de travail adoptée pour la réalisation du glossaire est décrite dans Preite, Zanella (2023). Le glossaire est accessible à l'url <https://www.lexi-term.unimore.it/glossarioptem/>

³ Nous envisageons à présent d'enrichir le glossaire avec l'allemand. D'autres langues pourront être ajoutées dans l'avenir, ainsi que d'autres produits de la région. Nous avons également en projet de transformer le glossaire en une base de données terminologique *machine readable*, avec une conséquente migration vers un site plus adapté.

⁴ « Alto contenuto territoriale », selon les mots de Asero, Patti (2009, p. 639).

Cependant à l'époque de la mondialisation, du marché international et surtout du tourisme culinaire qui se fonde sur la possibilité de conjuguer l'expérience touristique à la dégustation des produits typiques d'un terroir (Chessa, De Giovanni 2021), les produits locaux et les traditions alimentaires dépassent leurs frontières locales et nationales pour circuler dans d'autres communautés aux langues et aux cultures différentes, qui ne s'y identifient pas, mais qui en subissent de plus en plus la fascination⁵. En effet, comme le souligne Devilla (2015, p. 219), « [d]ans la construction de l'image d'une destination, la dimension gourmande constitue un élément important d'identification des lieux ». Et avec les produits de ces lieux, circulent également les noms qui les désignent, qui « traversent les frontières et sont intégrés dans un nombre élevé de langues du monde » (Kocoureck 1982, p. 134).

Qu'en est-il donc de ces dénominations lorsqu'elles sortent des frontières de la langue italienne ? C'est à cette question que nous essayons de répondre dans cet article. Après une présentation rapide du glossaire *Sapori locali* (section 2), nous allons comparer la forme des dénominations protégées dans la langue d'origine et celle des équivalents diffusés en français et en anglais (section 3). Cela nous permettra de relever l'influence que la valeur culturo-spécifique territoriale joue sur leur réception internationale.

2. *Sapori locali*

2.1. *Sélection de la nomenclature et des équivalents*

Actuellement le glossaire *Sapori locali* compte 53 entrées pour chacune des langues impliquées, qui correspondent à autant de produits émilien-romagnols jouissant de dénominations D.O.P., D.O.C., D.O.C.G., I.G.P. et I.G.T., sélectionnées dans la liste des dénominations italiennes inscrites dans le *Registre des appellations d'origine protégées, des indications géographiques protégées et des spécialités traditionnelles garanties (Règlement de l'Union Européenne n. 1151/2012 du Parlement européen et du Conseil du 21 novembre 2011 mise à jour 9 juillet 2024)*⁶.

⁵ Selon Chessa, De Giovanni (2021, p. 19) « Ciò implica la consapevolezza che territorio e prodotti siano indissolubilmente legati e che la scoperta dell'uno comporti la conoscenza degli altri, ma soprattutto che alla base vi sia una forte motivazione a scoprire l'uno per mezzo dell'altro ».

⁶ <https://www.politicheagricole.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/2090#main>

À cette liste s'ajoutent celles des vins de la Région :

<http://catalogoviti.politicheagricole.it/dopigp.phpet>

<http://catalogoviti.politicheagricole.it/denominazioni.php?codice=1058>

Pour ce qui est du repérage des équivalents des dénominations tels que diffusés dans les autres langues, nous adaptons à nos propos certaines démarches de la Terminologie textuelle qui prône la prise en charge de la réalité des usages tels qu'ils apparaissent dans les textes spécialisés, et de la situation langagière qui détermine certaines manifestations selon ses exigences : les termes sont donc à considérer comme des « éléments sémiotiques utilisés dans des situations de connaissances particulières » (Condamines 2018, p. 96). Ce qui se traduit dans la recherche des termes (à savoir des dénominations) dans des textes spécialisés dans le domaine œnogastronomique, ainsi qu'à adopter des instruments numériques pour leur traitement⁷. Ainsi, nous avons mené une recherche des équivalents par lesquels les langues cibles empruntent, traduisent ou calquent les dénominations italiennes, par la consultation de textes divers consacrés au domaine : d'une part, des sources primaires offrant des définitions, comme des bases de données fiables, dictionnaires et glossaires ; de l'autre part, des contenus apparaissant dans des sites de producteurs et de revendeurs (souvent multilingues), des sites officiels de l'Union européenne, etc.

Nous avons essentiellement fouillé le web partant des termes comme clés pour le repérage des textes, démarche ayant le mérite d'en montrer l'emploi en contexte pour chaque langue, ainsi que de permettre de découvrir les variantes⁸ diffusées des dénominations. En effet, nous avons pu remarquer que, sauf dans les réglementations juridiques, dans les textes en langues cibles les dénominations enregistrées sont généralement accompagnées de variantes diffusées chez les natifs. La prise en charge des variantes connues et employées par les usagers rend compte également de « [l]a circulation des termes [...] envisagée sous l'angle de la diversité de leurs usages sociaux » (Gaudin 2005, p. 81), selon les tenants de la Socioterminologie qui n'est pas du tout incompatible avec la Terminologie textuelle. Nous faisons l'hypothèse que cela permette de mieux identifier et construire la connaissance du produit.

2.2. Les fiches terminologiques

Chaque dénomination est traitée comme un terme auquel correspond une fiche terminologique composée de plusieurs champs : le domaine et le sous-

⁷ « [T]he specificity of textual terminology lies in the following: (i) it concerns only specialized corpora; (ii) the purposed of the study is considered throughout the analysis; (iii) terms are the point of entry into the texts; and (iv) tools are used as much as possible, with particular attention to the nature of the results obtained » (Condamines, Picton 2022, p. 235).

⁸ En cas de variantes dans la langue d'origine, la dénomination protégée officielle représente l'entrée et les variantes sont offertes dans un autre champ consacré à ce propos. En cas de variantes dans les langues cibles, c'est la plus fréquente qui est proposée comme entrée, alors que les autres sont énumérées dans le champ des variantes.

domaine, la marque grammaticale, la définition, la note, le contexte, la transcription phonétique, les variantes, l'étymologie, l'image du produit, les équivalents dans les différentes langues impliquées, les notes des équivalents et les sources bibliographiques et sitographiques consultées⁹, comme le montre, par exemple, la fiche consacrée à « Ciliegia di Vignola » (Fig. 1) :


Lemma	Ciliegia di Vignola	Immagine del prodotto	
Dominio e Sottodominio	Alimentazione – prodotto ortofrutticolo	Equivalenti	ENG Vignola Cherry ESP Cereza de Vignola FRA Cerise de Vignola RO Cilișgia di Vignola
Etichetta grammaticale	S f., sing	Nota degli equivalenti	Viene tradotto il nome del prodotto e rimane invariato il suo luogo di produzione.
Definizione	Frutto fresco dal sapore dolce e dalla polpa croccante, ottenuto dalle seguenti cultivar di ciliegio: Bigarneau Moreau, Mora di Vignola (precoci), Durone dell'Anella, Anellino, Gorgia, Durone Nero I, Sarrba, Van (medie), DuroneII, Durone della Marca, Lapini, Ferrovia, Sweet Heart (tardive). (1)	Ente	Università degli Studi di Modena e Reggio Emilia, Dipartimento di Studi Linguistici e Culturali
Nota	La "Ciliegia di Vignola IGP" è caratterizzata da una polpa consistente e croccante (latta eccezione per la Mora di Vignola) e da una buccia di un colore che va dal rosso brillante al rosso scuro (tranne che per la varietà Durone della Marca, dove invece è di colorogiallo e rosso brillante). (2) Frutti golosi e provvisti di proprietà nutrizionali, le ciliegie di Vignola hanno una polpa ricca di sali minerali come potassio, calcio, ferro, vitamina A e C e risultano essere diuretiche, rinfrescanti, astringenti. La zona di coltivazione si estende tra le province di Modena e Bologna e comprende, oltre a Vignola, 26 comuni delle zone pedicollinari. I frutti del ciliegio dolce maturano a primavera inoltrata, la raccolta si protrae da fine maggio a tutto giugno. (3) La Ciliegia di Vignola IGP va conservata in frigorifero, oppure in luogo fresco e asciutto. È il frutto ideale con cui chiudere il pasto e risulta un ottimo ingrediente per svariate ricette: dolci e salate: la "ciliegata", un dolce cotto nel vino e nello zucchero; le marmellate; il classico dolce "ciabuffi", salse per condire la caccagione, liquori, come il kirsch o lo cherry. Perfetta anche per la preparazione della frutta candita o sotto spirito. (4) Rappresentando il fiore all'occhiello dell'agricoltura locale, alle ciliegie di Vignola sono dedicati numerosi eventi fieristici ed enogastronomici. Si pensi, ad esempio, alla "Festa dei Ciliegi in Fiore", che dura tre settimane ed è famosa per i caratteristici carri allegorici. Questa manifestazione si celebra sin dagli anni Settanta. Si pensi anche a "Vignola è tempo di Ciliegia", un evento culturale e gastronomico molto atteso che fu istituito nel 1996: esso è caratterizzato dalla presenza di numerosi stand gastronomici che propongono una vasta gamma di pietanze a base di ciliegie, anche assai originali: risotti, liquori (kirsch, cherry brandy e il natafà), gelati e le classiche confetture. (5)	Data della scheda	23/03/2015
Contesto	Oltre alle peculiarità pedoclimatiche del territorio e all'eccezionalità del microclima sopra descritto, gli altri fattori che determinano l'eccezionale qualità e la reputazione della Ciliegia di Vignola sono la sapienza e la capacità dei produttori: queste vengono tramandate da padre in figlio attraverso le generazioni e consistono nella tecnica agronomica, nella raccolta e nel confezionamento del prodotto, effettuati esclusivamente a mano, che permettono di presentare al consumatore un prodotto unico nella sua specie. (6)	Ultima data di modifica	19/04/2023
Trascrizione fonetica	[ʧiˈlʲeʤja viˈɲoːlja]	Autore Revisore	Lucia Palazzese Chiara Buschi Gloria Zanella
Sinonimi/Varianti	Mora/Moretta di Vignola, Durone di Vignola: utilizzate come sinonimo di "ciliegia di Vignola", corrispondono ai frutti ottenuti da due varietà di ciliegio, rispettivamente mora di Vignola e Durone.	Fonti	1) http://agricoltura.regione.emilia-romagna.it/produzioni-agroalimentari/doc/prodotti-dop-e-igp-dellemilia-rom 2) http://agricoltura.regione.emilia-romagna.it/produzioni-agroalimentari/doc/prodotti-dop-e-igp-dellemilia-rom 3) http://www.italianbotanicalheritage.com/it/scheda.php?struttura=1426 4) http://agricoltura.regione.emilia-romagna.it/produzioni-agroalimentari/doc/prodotti-dop-e-igp-dellemilia-rom 5) https://www.italianfoodexperience.it/prodotti/ciliegie-vignola-piccole-sfere-rosse/ 6) http://www.ciboitaliano.com/dettaglio_ortofrutticoli.php?id=387 7) https://www.google.com/search?q=etimologia+ciliegia&rlz=C1FKPE_ah1T93217332&q=etimologia+ciliegia&app=chrome..69167j0j0z2330j0i10z1 8) https://www.italianfoodexperience.it/prodotti/ciliegie-vignola-piccole-sfere-rosse/ 9) http://eur-lex.europa.eu/LexUriServ.LexUriServ.do?uri=OJ:C:2012:052:0017:0021:FR:PDF
Etimologia	Ciliegia: dal lat. volg. *coresia, class. <i>corisium</i> ; dal gr. <i>keraasid</i> ; seconda metà sec. XIII. (7) Vignola: è un comune della circoscrizione di Modena in Emilia-Romagna.		

Figure 1
Ciliegia di Vignola. <https://www.lexi-term.unimore.it/ciliegia-vignola/>.

Chaque fiche présente les équivalents dans les autres langues (champ cerclé dans les Fig. 1, 2 et 3), qui sont aussi des liens actifs permettant d'accéder à la fiche correspondante dans la langue choisie où l'on retrouve les mêmes contenus (y compris définitions, descriptions, notes, etc.) que dans la langue d'origine. Ce système de renvois hypertextuels circulaires permet à l'utilisateur, pour chaque produit, de se déplacer parmi les différentes langues. À titre d'exemple, nous présentons également la fiche en français (Fig. 2) et en anglais (Fig. 3) :

⁹ D'autres champs renseignent sur l'organisation universitaire qui a rédigé et publié la fiche, sur la date de création de la fiche et sa dernière révision, sur l'auteur et le réviseur.

donne l'objectif d'« ordonner le réel ».

À la suite du contact entre langues et cultures, surtout culinaires, les produits dépassent leurs frontières géographiques et linguistiques, avec leurs dénominations qui circulent et s'adaptent aux nouvelles conditions d'usage. C'est la raison pour laquelle, pour chaque produit, nous avons recherché (cf. section 2.2) les équivalents déjà connus et répandus à l'étranger, qu'il s'agisse d'emprunts, de traductions, de calques ou d'autres termes sous influence (cf. Sablayrolles 2016)¹⁰. Avant d'étudier les stratégies à la base de leur passage en français et en anglais, nous allons examiner les dénominations italiennes selon leur forme.

3.1. Les formes des dénominations italiennes

Dans des études sur la sémantique des noms de produits typiques de la Sardaigne et de dénominations des produits fromagers italiens et français, Chessa, De Giovanni (2022, 2024) suggèrent qu'ils peuvent être constitués d'unités simples, ou monolexicales, ou bien complexes, ou polylexicales, aux structures syntaxiques différentes. Sur 53 dénominations que nous avons étudiées, seulement 5 sont monolexicales : *Brisighella*, *Ciccioli*, *Gutturnio*, *Lambrusco* et *Ortrugo*. Il est vrai que cela ne représente qu'un échantillon de l'ensemble des dénominations régionales de qualité, mais cette donnée semble significative du pourcentage réduit de termes simples (9,4%).

Pour le reste, il s'agit de dénominations polylexicales qui, à l'exception de *Cappello del Prete*, se composent d'une association binaire *produit + ville* (par exemple, « Coppa di Parma »), mais aussi *région* (« Pera dell'Emilia-Romagna »), *village* (par exemple, « Bianco di Castelfranco Emilia »), *coll/colline* (par exemple, « Colli Bolognesi Classico Pignoletto »), *fleuve* (par exemple, « Riso del Delta del Po »), *aire géographique* (par exemple, « Grana Padano »), et d'autres toponymes.

Notons que les dénominations complexes peuvent être réduites aux compositions syntagmatiques nominales suivantes, que nous reprendrons pour l'analyse des équivalents (section 3.2.1) :

- Nom + préposition + Nom géographique (par exemple, « Aglio di Voghiera », 18 items)
- Nom + Adjectif géographique (par exemple, « Coppia Ferrarese », 9 items)
- Nom + Nom géographique (par exemple, « Salame Felino », 6 items)

¹⁰ Tsai (2022, p. 216) va plus loin dans le classement des lexies de l'art culinaire sous l'influence d'autres langues et distingue « trois types d'emprunts lexicaux (emprunt intégral, emprunt adapté et emprunt modifié), le calque, l'hybride et l'allogénisme ».

- Nom + adjectif + préposition + Nom géographique (par exemple, « Asparago verde di Altedo », 6 items)
- NomGéographique + Nom (par exemple, « Colli di Parma Malvasia », 5 items)
- Nom géographique (par exemple, « Val Tidone », 3 items)

Indépendamment de la forme, 47 dénominations sur 53 (87%) montrent une indication géographique « [d]e sorte que, chez le natif, la mobilisation du nom de produit entraîne la mobilisation simultanée du nom du lieu réputé par ce produit » (Galissou 1991, p. 136). Ce type de dénomination devient donc un indicateur d'une spécialité œnogastronomique italienne réputée par le biais duquel le produit typique promeut le terroir et ce dernier à son tour valorise le produit.

3.2. Les équivalents français et anglais

Comme nous l'avons dit (section 1), le contact entre cultures permet également le partage de connaissances liées à l'offre œnogastronomique et, par conséquent, celui des noms respectifs des produits qui passent d'une langue à l'autre, en assumant des formes différentes. Une remarque préliminaire s'impose : étant bien entendu que toutes les dénominations enregistrées se trouvent telles quelles dans les sources des langues cibles consultées, dans leur rôle de *nom de produit*, c'est sur les variantes que l'on relève dans les mêmes sources que nous allons nous arrêter. Elles sont particulièrement intéressantes, parce qu'elles montrent sous quels noms ces produits sont connus par les publics étrangers. Lorsque présents, ces noms pourraient être considérés comme des sortes de reformulations qui facilitent la (re)connaissance du produit.

3.2.1. Les produits liquides : vins, huiles et vinaigres

Nous commençons notre analyse par les dénominations¹¹ de vins, huiles et vinaigres. Comme le rappelle Kleiber (2019, p. 7) les vins se distinguent par « les cépages et l'origine géographique [...] dont les noms servent métonymiquement de nom au vin qui y est produit ». Sauf pour les cépages, il en va de même pour les huiles et les vinaigres.

En ce qui concerne les unités monolexicales « Brisighella », « Gutturino », « Lambrusco » et « Ortrugo », auxquelles nous assimilons « Colline di Romagna », « Val Tidone » et « Terre di Veleja », noms réduits au seul toponyme (d'ailleurs aussi « Brisighella » est un toponyme sans d'autres précisions), il convient de constater que, dans les sources consultées

¹¹ Dans le monde viticole on parle d'*appellations*.

– qu'il s'agisse de cépages ou d'origines géographiques – les *liquides* gardent leur nom de produit enregistré inchangé : il s'agit donc d'emprunts. Cependant, les mêmes sources montrent également que ces dénominations peuvent être accompagnées par la précision fr. *huile/ vin* et ang. *oil/ wine*, ce qui inclut les produits dans leurs hyperonymes et en facilite la reconnaissance. En anglais, les toponymes antéposés à l'hyperonyme assument la valeur de déterminant : ang. « Brisighella/ Colline di Romagna *oil* » ; « Gutturino/ Lambrusco/ Ortrugo/ Terre di Veleja/ Val Tidone *wine* ».

Pour ce qui est des compositions syntagmatiques nominales montrant une origine géographique, trois structures possibles concernent les dénominations des liquides. Si les deux dernières concernent également les dénominations des nourritures, la première est exclusive des vins.

- Nom Géographique + Nom (+ Nom géographique)

Italien	Français	Anglais
Colli Bolognesi Classico Pignoletto	(Vin) Colli Bolognesi Classico Pignoletto	Colli Bolognesi Classico Pignoletto (wine)
Colli di Parma Malvasia	(Vin) Colli di Parma Malvasia	Colli di Parma Malvasia (wine)
Colli di Parma Rosso	(Vin) Colli di Parma Rosso	Colli di Parma Rosso (Red wine)
Colli Piacentini Monterosso Val d'Arda	(Vin) Colli de Piacenza Monterosso Val d'Arda	Colli from Piacenza Monterosso Val d'Arda (wine)
Colli Piacentini Trebbianino val Trebbia	(Vin) Colli de Piacenza Trebbianino Val Trebbia	Colli from Piacenza Trebbianino Val Trebbia (wine)

Pour commencer, une structure exclusive de certaines dénominations de vins montre l'origine géographique qui précède le nom du cépage. L'hyperonyme fr. *vin/* ang. *wine* apparaît dans certaines sources. Dans les deux derniers exemples, on ajoute une deuxième indication géographique qui concerne la vallée des mêmes cols et l'adjectif rétablit la forme *préposition + Nom géographique* en français et en anglais (« Colli *de/from* Piacenza »).

- Nom + préposition + Nom Géographique

Italien	Français	Anglais
Bianco di Castelfranco Emilia	(Vin) Bianco di Castelfranco Emilia	Bianco di (White wine from) Castelfranco Emilia
Bianco di Scandiano	(Vin) Bianco di Scandiano	Bianco di (White wine from/of) Scandiano
Fortana del Taro	(Vin) Fortana del / du Taro	Fortana del Taro (wine)

Dans ce cas, les noms des cépages et les toponymes qui suivent sont empruntés tels quels, mais nous avons relevé des variantes avec la traduction

du joncteur prépositionnel (par exemple, fr. « Fortana *du* Taro »), avec concurrence entre *from* et *of* pour l'anglais, ainsi que des variantes traduisant aussi *Bianco* en *White wine*. L'hyperonyme fr. *vin*/ ang. *wine* apparaît dans certaines sources.

- Nom + adjectif + préposition + Nom Géographique

Italien	Français	Anglais
Aceto balsamico di Modena	Vinaigre balsamique de Modène	Balsamic Vinegar from Modena
Aceto balsamico tradizionale di Modena	Vinaigre balsamique traditionnel de Modène	Traditional Balsamic Vinegar from Modena
Aceto balsamico tradizionale di Reggio Emilia	Vinaigre balsamique traditionnel de Reggio Emilia	Traditional Balsamic Vinegar from Reggio Emilia

En plus du toponyme, qui peut montrer une variante lorsqu'elle existe dans la langue étrangère (*Modène*) et de la traduction du joncteur prépositionnel, les dénominations des vinaigres (it. *aceto*/ ang. *vinegar*) se caractérisent par une double classification opérée par l'adjectif it. *balsamico*/ fr. *balsamique*/ ang. *balsamic* qui les différencie par rapport à d'autres types de vinaigre (par exemple, de vin ou de cidre), mais aussi, dans les deux derniers cas, par l'adjectif it. *tradizionale*/ fr. *traditionnel*/ ang. *traditional*, qui indique une catégorie plus fine¹². Les dénominations résultent être la traduction par transcription du modèle italien dans l'ordre morphosyntaxique des langues cibles, selon la stratégie du « calque morphologique de type littéral adapté sur le modèle de la composition syntagmatique nominale », que Di Spaldro, Auger et Ladouceur (2010, p. 164) appellent « technoscientifique », et que nous adoptons après avoir soutenu l'assimilation de ces dénominations à la terminologie technique.

3.2.2. Les nourritures

Pour ce qui est des dénominations polylexicales des nourritures, nous allons les présenter par structures compositionnelles avec leurs équivalents et nous allons ensuite observer quelles transformations elles subissent dans le transfert vers le français et l'anglais.

- Nom + préposition + Nom Géographique (15 items)

Italien	Français	Anglais
Aglio di Voghiera	Ail (de) Voghiera	Voghiera Garlic
Ciliegia di Vignola	Ciliegia di (cerise de)	Vignola Cherry

¹² Sur les notions de *tradizionale* et de *tipico* dans le cadre des textes normatifs italiens et de l'Union européenne, voir Biasiotti, Strambi (2015).

	Vignola	
Coppa di Parma	Coppa di Parma	Coppa from Parma
Culatello di Zibello	Culatello di Zibello	Culatello from Zibello
Formaggio di Fossa di Sogliano	Formaggio Fossa Sogliano Fromage de Fossa di Sogliano	Fossa Sogliano Cheese
Fungo di Borgotaro	Champignon de Borgotaro	Borgotaro Mushroom
Marrone di Castel del Rio	Marrone (de) Castel Rio	Chestnut from Castel Rio
Patata di Bologna	Pomme de terre Bologne	Potato from Bologna
Pera dell'Emilia-Romagna	Poire Emilia-Romagna	Emilia Romagna Pear
Pesca e nettarina di Romagna	Pêche et Nectarine de Romagne	Romagna Peach and Nectarine
Prosciutto di Modena	Jambon de Modène	Ham from Modena
Prosciutto di Parma	Jambon de Parme	Ham from Parma
Riso del Delta del Po	Riz du Delta du Pô	Po Delta Rice
Scalognone di Romagna	Échalote Romagne	Romagna shallot
Squacquerone di Romagna	Squacquerone de Romagne	Squacquerone Romagna

• Nom + adjectif + préposition + Nom Géographique (3 items)

Italien	Français	Anglais
Amarene brusche di Modena	Amarene Brusche di Modena / Griottes de Modène	Sour cherries from Modena
Asparago verde di Altedo	Asperge verte d'Altedo	Green asparagus of Altedo
Vitellone Bianco dell'Appennino Centrale	Taurillon blanc de l'Apennin Central	Vitellone Bianco from Appennino centrale

• Nom + Nom Géographique (6 items)

Italien	Français	Anglais
Cotechino Modena	Cotechino Modena	Cotechino (from) Modena
Mortadella Bologna	Mortadelle Bologne	Mortadella Bologna
Provolone Valpadana	Provolone Valpadana	Provolone Valpadana cheese
Salame Cremona	Saucisson Cremona	Cremona Salami
Salame Felino	Saucisson Felino	Felino Salami
Zampone Modena	Zampone Modena	Zampone (from) Modena

• Nom + Adjectif Géographique (9 items)

Italien	Français	Anglais
Coppa Piacentina	Coppa Piacentina	Coppa from Piacenza
Coppia Ferrarese	Coppia Ferrarese	Coppia Ferrarese
Grana Padano	Grana Padano	Grana Padano
Melone Mantovano	Melon mantovano	Mantua Melon
Pancetta Piacentina	Lard de Plaisance	Bacon from Piacenza
Parmigiano Reggiano	Parmesan	Parmigiano Reggiano
Piadina Romagnola	Piadina Romagnola	Piadina from Romagna

Salame Piacentino	Saucisson de Plaisance	Salame from Piacenza
Salamini Italiani alla Cacciatora	Saucisson chasseur	Italian salami cacciatora

Les stratégies qu'il est possible d'observer dans le passage de l'italien vers le français et l'anglais sont essentiellement trois, mais elles n'ont pas la même fréquence.

1. « Ciccicoli » (la seule unité monolexicale en dehors des vins/huiles) et « Cappello del Prete » (la seule unité polylexicale dépourvue de localisation géographique) sont traduits par un homologue fonctionnel¹³ en français *Rillons* et *Macreuse*, mais par une sorte de description des découpes de viande en question en anglais *Pork cracklings* et *Beef clod*.
2. Dans certains cas le français emprunte la dénomination enregistrée, indépendamment de la morphologie compositionnelle du terme : « Coppa di Parma », « Culatello di Zibello », « Cotechino Modena », « Provolone Valpadana », « Zampone Modena », « Coppa Piacentina », « Coppia Ferrarese », « Grana Padano » et « Piadina romagnola ». Les emprunts à l'identique en anglais concernent seulement « Coppia Ferrarese » et « Grana Padano ».
3. Pour le reste, il s'agit essentiellement d'emprunts de construction, c'est-à-dire de variantes obtenues par calque morphologique (comme pour les vinaigres) de différentes manières. Pour commencer, l'anglais montre une tendance majeure que le français à garder ou rétablir la forme avec le joncteur prépositionnel ang. *from/ of*, fr. *de*. Et cela surtout s'il s'agit d'emprunts dans lesquels on ne traduit que la préposition : ang. « Coppa *from* Parma/ Piacenza », « Culatello *from* Zibello », « Cotechino/ Zampone *from* Modena », « Coppa *from* Piacenza », « Vitellone Bianco *from* Appennino Centrale », etc., alors que le français garde la forme d'origine, sauf pour « Marrone *de* Castel Rio » (mais l'on perd dans les deux langues la préposition du toponyme, Castel *del* Rio).

S'il s'agit de calques qui traduisent aussi le nom du produit (mais parfois gardent inchangé le toponyme) les deux langues font des choix semblables mais pas toujours identiques : ang. « Ham from Parma »/ fr. « Jambon de Parme », ang. « Sour cherries from Modena »/ fr. « Griottes de Modène », ang. « Green Asparagus of Altedo »/ fr. « Asperge Verte d'Altedo », ang. « Bacon from Piacenza »/ fr. « Lard de Plaisance », etc.

Des cas contraires existent, cependant, dans lesquels l'anglais et/ou le français éliminent le joncteur et transforment une dénomination syntagmatique en une juxtaposition synthétique : ang. « Borgotaro

¹³ Parfois la traduction des *realia* se fait par remplacement avec un homologue local ou fonctionnel du phénomène de la culture source (Osimo 2011).

Mushroom » mais fr. « Champignon *de* Borgotaro », ang. « Squacquerone Romagna » mais fr. « Squacquerone *de* Romagna », ou bien « Pomme de Terre Bologne » mais ang. « Potato *from* Bologne », etc. En tout cas la réduction de la dénomination italienne reste assez fréquente : fr. « Échalote Romagne », « Poire Emilia-Romagna », « Ail (de) Voghiera », « Formaggio Fossa (de) Sogliano », ang. « Voghiera Garlic », « Vignola Cherry », « Fossa Sogliano Cheese », « Emilia Romagna Pear », « Po Delta Rice », « Romagna Shallot ». Remarquons que ces dénominations binaires rétablissent l'ordre morphosyntaxique de l'anglais, avec le nom de lieu qui assume la fonction adjectivale de déterminant, ce qui n'est pas le cas pour les dénominations syntagmatiques analytiques avec joncteur prépositionnel.

Enfin, il y a également des dénominations qui montrent une construction synthétique en italien dans lesquelles le Nom du produit est modifié soit par un Nom soit par un Adjectif géographique. La réaction des langues d'arrivée est multiple. Nous retrouvons la tendance au rétablissement de la préposition (par exemple, ang. « Salame *from* Piacenza »/ fr. « Saucisson *de* Plaisance ») et de l'ordre morphosyntaxique de l'anglais (« Mantua Melon », « Felino Salami », « Provolone Valpadana (cheese) », « Italian salami cacciatore »), mais aussi des emprunts tels quels (ang. « Mortadella Bologna », « Parmigiano Reggiano », fr. « Cotechino Modena », « Piadina romagnola », « Grana Padano », etc.). Parfois on traduit le nom du produit : ang. « Cremona Salami », « Bacon from Piacenza », fr. « Parmesan », « Saucisson Felino », « Mortadelle Bologna », « Saucisson de Plaisance ». Pour terminer, il arrive que pour la construction Nom + Adjectif géographique, ce dernier est reconduit au Nom propre : ang. « Coppa/Bacon from Piacenza » (*piacentina*), « Piadina from Romagna » (*romagnola*), « Mantua melon » (*mantovano*), fr. « Lard/ Saucisson de Plaisance » (*piacentina*).

3. Conclusions

Si nous considérons que les dénominations protégées appartiennent à la terminologie technique, nous acceptons également que les variantes dans les langues cibles sont des cas de formation terminologique secondaire (Sager 1997), car elles sont influencées par l'existence d'un terme dans la langue source. Ainsi, un savoir passe d'une communauté linguistique à une autre, par le biais d'adaptations et de transformations légères¹⁴ (traductions, emprunts et

¹⁴ D'ailleurs, comme le notait Cartier (2019), pour des raisons culturelles, un grand nombre de plats étrangers aussi gardent leur nom d'origine avec peu d'adaptations.

calques) qui permettent aux termes de se diffuser et de s'installer dans les langues qui les reçoivent. En tout cas, le terme primaire, soit la dénomination protégée, apparaît normalement avec ses variantes dans les textes des langues cibles que nous avons consultés, ce qui permet à l'utilisateur, au touriste, à l'amateur, mais aussi à l'expert, de reconnaître un produit tout en bâtissant un pont entre connaissance technique et connaissance générale d'un produit.

Pour ce qui est de l'étude des dénominations protégées, une donnée qui ressortit visiblement et le lien étroit qui se noue entre un produit et l'endroit géographique de production. En effet, rappelons que sur 53 dénominations qui composent à présent le glossaire, seulement 5 ne contiennent pas de toponymes : « Ciccicoli », « Gutturnio », « Lambrusco », « Ortrugo » et « Cappello del Prete ». Pour le reste, il s'agit de dénominations polylexicales qui (à l'exception de l'unité monolexicale « Brisighella ») associent, selon des schémas différents (section 3.1), un Nom de produit et un toponyme. Il s'ensuit que si le produit typique promeut le terroir, ce dernier valorise en retour le produit.

Dans le passage de la langue italienne au français et à l'anglais, si l'on veut faire référence au produit en tant qu'élément enregistré et protégé, les textes que nous avons consultés montrent l'emprunt de la dénomination d'origine. Cependant, ce n'est pas (ou pas seulement) celle-ci qui circule parmi les amateurs, les touristes, le grand public en général. C'est la raison pour laquelle les textes consacrés à la diffusion dans le web des connaissances des produits protégés ne se limitent pas à les présenter par leur dénomination officielle, mais utilisent aussi des variantes plus populaires. Ce qui permet de découvrir sous quels noms ces produits sont connus par les publics étrangers. Ainsi, parmi les liquides (section 3.2.1), les dénominations des huiles et des vins circulent essentiellement en forme d'emprunt de l'italien (éventuellement précisé par l'hyperonyme fr. *vin/ huile*, ang. *oil/ wine* qui permet l'identification de la catégorie d'appartenance), les vinaigres sont par contre des calques morphologiques de type littéral. Pour ce qui est des nourritures (section 3.2.2), les traductions et les emprunts sont rares : ce sont en effet les calques qui l'emportent. Mais ces calques, comme on a pu le remarquer, se réalisent par des stratégies multiples et parfois opposées : avec élimination ou rétablissement du joncteur prépositionnel, avec traduction de la seule préposition ou bien aussi du Nom de produit et/ou du Nom géographique, avec transformation de l'Adjectif géographique en Nom, etc.

Or dans un tel foisonnement dénomiatif, il est difficile de prévoir quelle pourrait être en anglais et en français l'équivalent correct (ou plus simplement le plus diffusé et fréquent) d'une dénomination italienne. C'est donc pour guider le public dans ce choix que nous avons construit *Sapori locali*, le glossaire multilingue des produits protégés et enregistrés de l'Émilie-Romagne.

Bionote : Chiara Preite est maître de conférences en Langue et traduction française à l'Université de Modène et Reggio d'Émilie (IT), membre du collège de Doctorat en Sciences Humaines et directrice de thèses en linguistique française. Elle coordonne les groupes de recherche *Modena Lexi-Term* et *Do.Ri.F Socioterminologie et textualité*. Ses axes de recherche principaux concernent le français juridique, la lexicographie – générale et spécialisée –, la terminologie, la vulgarisation des connaissances spécialisées. Depuis 2021 elle a obtenu l'habilitation nationale au rôle de Professeur des Universités.

Adresse de l'auteure : chiara.preite@unimore.it

Remerciements : nous remercions le Département d'Études Linguistiques et Culturelles de l'Université de Modène et Reggio d'Émilie pour le soutien économique pour la réalisation du glossaire *Sapori locali*.

Références bibliographiques

- Asero V. e Patti S. 2019, *Prodotti enogastronomici e territorio: la proposta dell'enoturismo*, in Becheri E. (a cura di), *XVI Rapporto sul turismo italiano*, Franco Angeli, Milano, pp. 637-649.
- Biasiotti M.A. e Strambi G. 2015, *Incertezza e ambiguità della terminologia nel diritto agroalimentare: il ruolo delle strategie semantic based*, in Chessa F., De Giovanni C. et Zanola M.-T. (a cura di), *La terminologia dell'agroalimentare*, Franco Angeli, Milano, pp. 67-83.
- Cadeddu J., Kauffer M. et Keromnes Y. (éds.) 2019, *La gastronomie à l'ère numérique*, Stauffenburg, Tübingen.
- Cartier E. 2019, *Emprunts en français contemporain : étude linguistique et statistique à partir de la plateforme Néoveille*, in Kacprzak A., Mudrochová R. et Sablayrolles J.-F. (éds.), *L'emprunt en question(s)*, Lambert Lucas, Limoges, pp. 145-186.
- Chessa F. e De Giovanni C. 2021, *Comunicare il patrimonio enogastronomico sardo attraverso Wikipedia*, in Devilla L. et Galiñanes Gallén M. (a cura di), *Lingue minori e turismo*, Arkadia, Cagliari, pp. 19-32.
- Chessa F. e De Giovanni C. 2022, *La denominazione del prodotto gastronomico nell'offerta turistica: breve analisi dei prodotti dolciari sardi*, in Devilla L. et Galiñanes Gallén M. (a cura di), *Le parole del turismo*, Edizioni dell'Orso, Alessandria.
- Chessa F. et De Giovanni C. 2024, *Les produits fromagers AOP et IGP français et italiens. Pour une sémantique de la dénomination appliquée à la terminologie*, in « Roczniki Humanistyczne » 8, pp. 35-47.
- Chessa F., De Giovanni C. e Zanola M.-T. (a cura di) 2014, *La terminologia dell'agroalimentare*, Franco Angeli, Milano.
- Condamines A. 2018, *Nouvelles perspectives pour la terminologie textuelle*, in Altmanova J., Centrella M. et Russo K.E. (éds.), *Terminology and Discourse / Terminologie et discours*, Peter Lang, Berne, pp. 93-112.
- Condamines A. and Picton A. 2022, *Textual Terminology. Origins, principles and new challenges*, in Faber P. and L'Homme M.-C. (eds.), *Theoretical Perspectives on Terminology. Explaining terms, concepts and specialized knowledge*, Benjamins, Amsterdam / Philadelphia, pp. 219-236.
- Devilla L. 2015, *Le rôle de la gastronomie dans la représentation de l'identité régionale sarde*, in Paissa P., Rigat F. et Vittoz M.-B. (éds.), *Dans l'amour des mots*, Dell'Orso, Alessandria, pp. 219-229.
- Di Spaldro J., Auger P. et Ladouceur J. 2010, *Le calque technoscientifique : un procédé néologique avantageux pour la terminologie française ?*, in « Néologica » 4, pp. 163-183.
- Fischler C. 2001, *L'omnivore*, Odile Jacob, Paris.
- Frøeliger N. 2021, *La terminologie, outil de vulgarisation et de mise en discours pour la traduction pragmatique*, in Vaupot et al. (éds.), *Contacts linguistiques, littéraires, culturels*, Presses universitaires de Ljubljana, Ljubljana, pp. 72-89.
- Galisson R. 1987, *Accéder à la culture partagée par l'entremise des mots à C.C.P.*, in « Études de Linguistique Appliquée » 67, pp. 119-140.
- Galisson R. 1988, *Culture et lexiculture partagées : les mots comme lieux d'observation des faits culturels*, in « Études de Linguistique Appliquée » 69, pp. 74-90.
- Galisson R. 1991, *De la langue à la culture par les mots*, Clé International, Paris.
- Gaudin F. 2005, *La socioterminologie*, in « Langages » 157, pp. 80-92.
- Kleiber G. 2019, *Le vin et ses déterminants*, in Cadeddu J., Kauffer M. et Keromnes Y.

- (éds.), *La gastronomie à l'ère numérique*, Stauffenburg, Tübingen, pp. 15-32.
- Kocoureck R. 1982, *La langue française de la technique et de la science*, Brandstetter, Wiesbaden.
- Lerat P. 2016, *Langue et technique*, Hermann, Paris.
- Marti A., Planelles Iváñez M. et Sandakova E. (éds.) 2018, *Langues, cultures et gastronomie : communication interculturelle et contrastes*, Peter Lang, Berlin.
- Osimo B. 2011, *Manuale del traduttore: guida pratica con glossario*, Hoepli, Milano.
- Preite C. et Zanella G. 2023, « *Sapori Locali* » : *Glossaire multilingue des produits typiques émiliens*, in Di Nunzio G.M., Costa R. et Vezzani F., *La terminologie numérique multilingue aujourd'hui*, <https://ceur-ws.org/Vol-3427/short6.pdf> (9.05.2024)
- Sablayrolles J.-F. 2016, *Emprunts et influences d'autres langues*, in Hildebrand Z., Kacprzak A. et Sablayrolles J.-F. (éds.), *Emprunts néologiques et équivalents autochtones en français, en polonais et en tchèque*, Lambert Lucas, Limoges, pp. 23-35.
- Sager J. 1997, *Term formation*, in Wright S.E. and Budin G. (eds.), *Handbook of terminology management*, vol. 1, pp. 25-41.
- Tsai C.-W. 2022, *Emprunts et influences d'autres langues dans le français de l'art culinaire*, in « Cahiers de lexicologie » 120 [1], pp. 211-235.
- Zanola M.-T. 2015, *La terminologia agroalimentare: normatività e comunicazione plurilingue*, in Bombi R. e Orides V. (a cura di), *Italiani nel mondo*, Editrice Universitaria Udinese, Udine, pp. 145-152.

QUELLE SOUPLESSE POUR LE SIGNE LINGUISTIQUE DU NOM DE MARQUE ? Étude de sa variation phraséologique au prisme de la créativité du discours numérique au quotidien

MICHELA TONTI
UNIVERSITÀ DEGLI STUDI DI BERGAMO

Abstract – In the present work, we propose a short diachronic study in which we aim to identify the cultural anchorage of the phraseological units identified in the lexical inventory of the blogosphere, while attempting to demonstrate that certain collocations that affect many areas of reality display a rough outline of semantic stability. The dynamism of brand names is highlighted because they represent our empirical base. The concept of collocation draws on multiple theoretical frameworks, namely the feature of quantitative salience of certain familiar words in situational contexts, the feature of informational salience emphasises the meaning that the brand name acquires in context. The use of digitized resources as well as the study of collocations in the Firthian sense adopted by corpus linguistics, can be a potentially powerful tool for analyzing brand names. Our aim is to study collocations of prototypical nouns that would deserve to be included in general language dictionaries for their cultural representativeness. We propose comparisons between the theory of the prototype, which as a theory of reference, implements a semantics of designation just like the dictionary, and our collocations whose base, the central notion, consists of a brand name.

Keywords: brand name; collocations; corpus linguistics; cultural implicit; dictionary.

1. Introduction

Un Nom de Marque (désormais NdM) correspond, dans sa dimension juridique, à un espace sémantique bien balisé. Dans le contexte de la communication publicitaire et du marketing, c'est un nom au pouvoir évocateur. Pour l'ensemble des locuteurs d'une langue il représente une structure langagière afférant à leur stock lexiculturel. Parmi ces locuteurs, nous nous intéresserons ici en particulier aux « écrilecteurs »¹ (Paveau 2017)

¹ Dans la perspective de l'analyse du discours numérique, on appellera écriture numérique une production scripturale impliquant des traits graphiques, langagiers et discursifs spécifiques dus à la conversion numérique et plus généralement une conception modifiée de la culture de l'écrit et plus généralement de la discursivité. [...] Les formes de l'écriture numérique sont marquées par la contrainte technique, présentent des caractéristiques discursives, énonciatives et sémiotiques

et à leur usage des NdM dans des blogs, pour la plupart personnels². Dans des études précédentes (Tonti 2019, 2020a, 2020b, 2021a, b et c), nous avons démontré qu'un NdM partage aussi bien les caractéristiques d'un nom propre que d'un nom commun³ lorsqu'il circule dans le discours au quotidien, dans cette étude, nous abordons ses mécanismes d'emploi au niveau sémantique et notamment phraséologique. Faisant écho aux découvertes les plus récentes sur l'organisation cognitive humaine (Tomasello 2019), nous adoptons une approche fonctionnelle, à savoir une approche socio-culturelle discursive sur corpus afin de mettre en valeur la créativité et les innovations langagières en matière d'unités multilexémiques – notamment des collocations issues d'un NdM. Notre approche culturelle pour le traitement de l'onomastique commerciale dépasse les frontières jusqu'ici balisées par les études existantes et d'expression française parmi lesquelles nous signalons celles de Berthelot-Guiet (2003), Fèvre-Pernet (2005), Laurent (2010) et Altmanova (2013, 2018, 2019).

Dans un premier temps, et d'un point de vue théorique, nous nous questionnerons sur la notion de collocation⁴ constituée d'un NdM dans une perspective lexicographique et dictionnaire. Robert Galisson et Jean-Claude André dans les années 1990 avaient établi leur *Dictionnaire des Noms de marque courants* en se fixant pour objectif de proposer un dictionnaire interstitiel en mesure de combler les lacunes des dictionnaires de la langue

(délinéarisation, augmentation, hybridation) et possèdent des caractéristiques discursivo-communicationnelles particulières (investigabilité, imprévisibilité, dissémination) (Paveau 2017, pp. 133-134).

² Les blogs ont incarné, dès le début des années 2000, l'entrée progressive de l'Internet dans les pratiques culturelles et médiatiques. L'internaute devient ainsi producteur de contenus mais deux réalités, a priori distinctes, se dessinent. Selon Trédan (2010), il faut envisager le blog, d'un côté, comme support communicationnel utilisé dans la gestion des affinités amicales et, de l'autre, comme support informationnel utilisé par les médias. L'internaute est saisi au prisme de son double statut : animateur d'un espace de publication et lecteur de sites d'informations généralistes ou thématiques.

³ Dans le cadre de la thèse de recherche doctorale (2019) et des recherches successives que nous avons conduites et notamment en 2020a et b, 2021a, b et c, 1987, les NdM en discours ont fait l'objet de notre exploration. Au fil de 1987 signes linguistiques portant sur des NdM, nous avons remarqué que les locuteurs peu soucieux des critères linguistiques, économiques et financiers, et pourtant désireux d'avoir recours à des signes linguistiques qui appartiennent à leur stock lexical, se servent des NdM comme de tout autre nom propre ou nom commun. Au sein de l'étude (2020a), nous signalons la section I.4.1. « Une miroitante instabilité d'équivalences : un Nom propre est-il un NdM comme un Nom commun est un Nom de produit ? » (pp. 35-39) qui passe en revue une sélection récente de la littérature existante sur l'onomastique commerciale avec une attention particulière aux aspects formels et descriptifs. Alors que le chapitre II (Tonti 2020a), « Pour une linguistique *in vivo* », signe l'acte de naissance de notre réservoir de NdM dont le potentiel est loin d'être épuisé du point de vue lexico-culturel et discursif.

⁴ Nous soulignons qu'une analyse minutieuse de la littérature existante portant sur le concept de collocation exigerait une place trop ample au sein d'un article de type expérimental orienté par corpus.

générale. La charge culturelle partagée par tout natif – ou ayant un vécu linguistique important – et l’implicite culturel sous-jacent à un NdM donné, ainsi que sa notoriété dans le discours au quotidien, constituaient les critères de sélection et d’intégration du NdM dans ledit dictionnaire. Ici, slogans, signatures et signes linguistiques confectionnés par les publicitaires étaient retenus, alors que, pour notre part, nous visons les discours au quotidien circulant autour des NdM et les unités multilexémiques issues de la créativité des locuteurs. Ces unités répondent aux exigences d’un dictionnaire de collocations. Nous puiserons dans le travail dirigé par Orlandi et Giacomini (2016), *Defining collocations for lexicographic purposes : from linguistic theory to lexicographic practice*. D’un point de vue méthodologique, il s’agira de confirmer l’intérêt des corpus pour toute linguistique située (Condamines, Narcy-Combes 2015) ; en l’occurrence, notre fouille est intégralement conduite à partir du corpus *frTenTen* (2017, 10 milliards de tokens) avec une approche guidée sur corpus et outillée. Les pistes de réflexion proposées par Orlandi dans son essai (2016), retraçant entre autres l’état de l’art des définitions fournies en littérature pour le concept de collocation, nous ont aiguillée dans notre recherche initiale de NdM saillants.

Nous nous proposons d’analyser une base empirique constituée essentiellement de cinq NdM : *Louboutin*, *Kodak*, *Chocapic*, *Canada Dry*, *Port Salut* qui font appel à des référents commerciaux et culturels fort diversifiés, respectivement : des chaussures, des appareils photos, des produits céréaliers fortement liés à la culture générationnelle de l’enfance et de l’adolescence et pour finir une boisson et un fromage. Cependant, nous allons principalement nous concentrer sur une étude de cas conduite sur le NdM *Canada Dry* qui constitue le socle dur des exemples analysés dans notre deuxième partie, alors que dans notre troisième partie, nous nous tournerons vers les NdM *Chocapic* et *Port Salut* qui produisent des « phrases situationnelles » (Lamiroy, Klein 2016).

Dans le présent travail, nous proposons donc une étude en diachronie courte dans laquelle nous poursuivons le but de relever l’ancrage culturel des unités phraséologiques repérées dans l’inventaire lexical de la blogosphère tout en essayant de démontrer que certaines collocations qui affectent maints domaines du réel, affichent une ébauche de stabilité sémantique. Précisons que notre domaine d’observation est expérimental ; l’angle de vue ciblé est fort spécifique et il permet tantôt d’accueillir certaines assises de base, tantôt de les modifier. Il en découle que le choix de certaines approches pour cerner des collocations culturellement saillantes constituées de NdM n’est pas généralisant.

2. En quête d'une définition de collocations issues de NdM à but lexicographique

2.1. Cadre théorique

Dans notre parcours heuristique, nous procédons par approches successives en éliminant progressivement les définitions de collocations qui ne s'adaptent pas à notre corpus et en ne conservant qu'une gamme restreinte de solutions tendant vers celle qui nous semble optimale. Nous présentons donc une sélection d'approches définitoires sans aucune prétention d'exhaustivité car un passage en revue minutieux du concept de collocation engagerait un nombre de pages disproportionné pour un article qui s'appuie sur des données empiriques. Lamiroy (2016) estime que la propriété définitoire de la collocation réside dans son caractère binaire (Mel'čuk 2013 ; Grossman, Tutin 2003). Reposant sur un phénomène d'attraction lexicale essentiellement statistique, elle est le résultat d'un processus diachronique à la suite duquel une combinaison de mots devient progressivement figée et son sens devient à son tour prototypique et donc opaque. De plus, lorsque le sens de la collocation est conventionnel, il n'est pas totalement transparent, même pour des natifs de la langue. De fait, nous nous proposons de cerner ici et de mettre en valeur des unités dont le constituant est un NdM, c'est-à-dire un « palimpseste verbo-culturel ». Nous recourons à cette dénomination galissonienne (Galisson 1993) pour indiquer le cas où l'implicite sous-jacent des unités phraséologiques est « le produit du chevauchement d'un sous-énoncé lexicalisé et d'un sur-énoncé résultant de la déconstruction (délexicalisation) du sous-énoncé de base » (Galisson 1993, p. 43). À titre d'exemple, nous mentionnons l'énoncé suivant : « Aujourd'hui, il reste des canada dry de services publics ça ressemble ce n'en est plus parce que les critères de rentabilité ont remplacé ceux d'efficacité » (blogs.fr, datation non disponible). Le décodage s'effectue par reconnaissance des propriétés caractéristiques de la boisson *Canada Dry* qui a le goût et la couleur de l'alcool sans en être, désignant donc une chose ou une personne qui a l'apparence de ce qu'elle prétend être sans en avoir les qualités. Les services publics sont donc assimilés aux faux airs de la boisson.

Fuertes-Olivera *et al.* (2012, p. 299) s'écartent des définitions précédemment mentionnées concernant le terme collocation :

Le terme collocation est choisi en tant que terme parapluie se rapportant à une combinaison de mots typiques du contexte linguistique en question pouvant être remployée dans une production textuelle ou dans la traduction assistée d'un texte. Les collocations sont composées de deux ou de plusieurs mots, elles ne peuvent pas constituer une phrase mais disposent d'une saillance informationnelle au bénéfice de ses potentiels usagers. (Fuertes-Olivera *et al.* 2012, p. 294 ; notre traduction)

La définition est ici étroitement liée à la notion de saillance (Fuertes-Oliveira, Nielsen 2011 ; Bothma, Tarp 2012), c'est-à-dire « la condition d'être directement liée au domaine, à la (aux) fonction(s) du dictionnaire, à la situation d'utilisation, et au niveau de compétence des utilisateurs prévus » (Fuertes-Oliveira *et al.* 2012, p. 294 ; notre traduction). Le trait de la saillance doit être retenu comme l'un des paramètres à la base de nos exemples en contexte.

Les collocations sont considérées non seulement comme le résultat de contraintes lexicales, mais aussi de « contraintes pragmatiques sur la sélection lexicale » selon Bartsch (2004, p. 178 ; notre traduction). Cette double contrainte peut devenir relativement stable lorsque la combinaison intervient de manière relativement stable en raison de la fréquence de ses occurrences dans la même constellation et sens (2004, p. 71). À cet effet, Bartsch trouve un compromis entre l'approche phraséologique et l'approche empirique basée sur la distribution et les fréquences.

Une autre définition fonctionnelle de la collocation se trouve chez Giacomini :

Dans un sens étroit, les collocations étaient définies comme des expressions idiomatiques de plusieurs mots [...], soumises à une compositionnalité, une substituabilité et une possibilité de modification restreintes, et identifiables par des tests idiomatiques standard. Au sens large, les collocations sont à considérer comme des combinaisons de mots familiers récurrents dans notre lexique mental [...], principalement associés à des scènes typiques, c'est-à-dire à des contextes situationnels. (Giacomini 2010, p. 1184, notre traduction)

Si nous adaptons cette dernière définition au NdM, elle paraît intéressante à plusieurs titres : certains signes linguistiques de NdM au vu de leur longévité et notoriété intègrent le stock lexiculturel des locuteurs, quittant ainsi leur statut de terme déposé. Il s'agit de lexies courantes qui activent un réseau d'informations partagées et sont souvent utilisées dans des discours peu contrôlés comme c'est le cas pour notre batterie d'exemples issue d'une fouille sur corpus. Force est de constater qu'une collocation a une saillance qualitative lorsqu'elle est immédiatement associée à une image mentale (Williams 2013, p. 91). La saillance qualitative et l'emploi en situation d'une unité multilexémique répondent avec cohérence aux caractéristiques d'un dictionnaire.

Une dernière approche définitionnelle de collocation utile à des fins lexicographiques serait à chercher dans la Théorie du prototype. Selon Kleiber (1990, p. 49) « le prototype n'est vraiment considéré comme le meilleur exemplaire que s'il apparaît comme celui qui est le plus fréquemment donné comme tel ». Le prototype fonctionne en l'occurrence comme point de référence cognitif (*cognitive reference point*, Rosch 1975) de nos catégories et systèmes de classification. Cela suppose bien sûr que le

locuteur ait la compétence nécessaire pour juger du degré de similarité de tels appariements prototypiques. Le prototype se présente comme une construction mentale issue d'opérations cognitives (telles que des activités de discrimination). L'important est de souligner qu'il ne s'agit plus d'exemplaires, d'instances ou de représentants réels de la catégorie, mais d'une entité abstraite, construite sur la base des propriétés typiques de la catégorie. À l'instar de Orlandi (2016), nous estimons que la fréquence peut être également considérée comme une caractéristique prototypique à retenir lors de la sélection des collocations. En guise de conclusion, la théorie contextualiste firthienne avance une définition de collocation intéressante lorsque l'on travaille en linguistique de corpus, car d'après cette théorie, la signification et l'emploi d'un mot peuvent être étudiés à partir de l'observation de son contexte linguistique, le co-texte, qui met en évidence les liens syntaxiques et sémantiques caractérisant le mot-clé et ses co-occurents.

À partir de ces différentes pistes de réflexion, nous appliquons le cadre définitoire du prototype comme représentation des propriétés typiques pour sélectionner l'éventail de NdM alors que nous empruntons à Fuertes-Olivera le trait de la saillance informationnelle d'une collocation dans une situation donnée et à Giacomini le trait de la récurrence de certains mots familiers qui viennent à l'esprit dans des contextes situationnels, à savoir l'enjeu de la saillance quantitative. L'analyse des *KWIC (Key Word In Context)* montre que des régularités sont observables, notamment un mot peut être décrit selon ses collocations, ses préférences sémantiques et à partir de son contexte linguistique (Di Vito 2013). Pour conclure, nous adaptons les traits définitoires de collocation proposés par Fuertes-Olivera (2012), Giacomini (2010) et le contextualisme firthien.

2.2. Cadre méthodologique

Au fil de nos études précédentes (Tonti 2020a et b, 2021a, b et c), nous avons proposé de dégager une perspective de recherche expérimentale qui puise ses racines dans des observations empiriques et de nature représentationnelle : la circulation et la prolifération créative du signe linguistique du NdM dans les pratiques discursives au quotidien sont mises à l'honneur. Certes, nous avons bien saisi le pouvoir du signe linguistique du NdM en tant qu'activateur lexiculturel, néanmoins, aucune de nos études précédentes ne portait son attention sur les unités phraséologiques. Le choix des NdM à explorer dans la présente étude a été orienté par l'effervescence des signes linguistiques *Canada dry*, *Port Salut* et *Chocapic* qui demeurent fort vivants dans la mémoire collective des locuteurs et dans leurs discours. La méthode d'investigation que nous allons utiliser pour mener notre recherche et répondre à la problématique posée est celle de l'approche guidée sur corpus

(*corpus-based*) (Tognini-Bonelli 2001). Notre recherche s'engage dans une conception relationnelle et fonctionnelle de la langue : nous postulons que ce sont les contextes d'usage qui vont faire émerger des sens, tour à tour variables, de nos noms pour lesquels nous allons rechercher des régularités et des variations d'usage au sein des unités phraséologiques qu'ils composent. Les exemples qui suivent sont tirés du corpus *frTenTen* (2017) qui comporte un large panel de supports textuels relevant d'extensions de domaines diversifiés d'origine nationale .fr, .be, .ch, .ca, des sites institutionnels, blogs personnels ou espaces blog réservés aux abonnés de la presse nationale, notamment française et mondiale. L'extraction des concordances dont le mot-clé est représenté par notre signe linguistique a été conduite dans un premier temps à l'aide d'expressions régulières. Cependant, afin de valoriser la plus ample palette de résultats, nous avons procédé par recherche simple (CQL)⁵ du mot tout en sélectionnant à la main les unités phraséologiques plus pertinentes au vu des critères d'analyse retenus.

3. La souplesse⁶ du NdM façonné par les locuteurs : *Canada dry, Port Salut, Chocapic*

3.1. *Canada dry* : le prototype comme représentation de propriétés typiques et ses collocations en discours

Selon Rey-Debove, « le dictionnaire est notre mémoire lexicale et notre sémantique lexicale de base ; il décrit du mieux qu'il peut les unités de la langue plus ou moins liées au référentiel » (1989, p. 143). Puisque notre but est d'étudier des collocations de noms prototypiques qui mériteraient de figurer dans des dictionnaires de langue générale pour leur représentativité culturelle, nous proposons des rapprochements entre la théorie du prototype qui en tant que théorie de la référence, met en œuvre une sémantique de la désignation tout comme le dictionnaire et nos collocations dont la base, la notion centrale, se compose d'un NdM.

Dans un premier temps, nous proposons d'analyser le NdM *Canada dry* qui convient pour sa valeur de prototype, car il fonctionne comme point de référence cognitif de nos catégories et systèmes de classification qui sont donc le fruit d'appariements prototypiques. Or, toute la symbolique sémiotique d'une société montre que la notion sociale de « *Canada dry* » a un

⁵ *Corpus Query Language*

⁶ Nous faisons appel au concept de « souplesse » car les mécanismes de la communication par Internet font émerger des phénomènes de *souplesse* qui se manifestent même dans la dimension référentielle traditionnellement attribuée aux noms propres. Les nombreux énoncés de notre article fournissent des exemples concrets de ce phénomène.

domaine d'application plus large que sa définition, c'est-à-dire un domaine d'application différent, déplacé, par rapport à la définition. En effet, le signe linguistique « Canada dry » circule rarement dans notre corpus *frTenTen* en tant qu'exemplaire ou représentant réel de sa catégorie commerciale correspondant à une boisson gazeuse, un soda aromatisé avec du gingembre qui se donne de faux airs de boisson alcoolisée en canette (21 occurrences). Il est plus fréquent en tant qu'entité abstraite, construite sur la base des propriétés typiques non pas de sa classe commerciale mais de sa manière d'être, qui est de communiquer une fausse ressemblance, de faux airs de quelque chose (369 occurrences). Cette apparence illusoire constitue le noyau qui assure la fixité centrale de la catégorie issue du NdM. Pour reprendre un exemple célèbre livré par Lakoff (1972) un moineau possède un degré d'appartenance plus grand à la catégorie d'oiseau qu'un pingouin ; à l'instar de cet exemple, *Canada dry* détient le palmarès en tant que micro-univers référentiel cohérent de la notion de ressemblance trompeuse et illusoire. Le lien associatif est assuré par le slogan qui a accompagné le signe linguistique du NdM, paradoxalement, d'autres NdM dont le référent est également représenté par des sodas pétillants et au parfum de gingembre comme *Schweppes Ginger Ale* (par exemple) ne parviennent pas à produire cet appariement prototypique. Au vu de la nature polysémiotique du signe linguistique NdM, ce dernier profite également des discours marketing et publicitaires construits autour de lui.

Nous avons analysé la distribution du NdM *Canada dry* (369 occurrences) et de ses collocations en contexte sur la base de la fonction syntaxique et donc du rôle que le signe linguistique « Canada dry » joue dans la structure d'ensemble de la phrase où il est employé, ce qui est préconisé par le concept de collocation issu de la théorie du contextualisme firthien. Lors de notre sélection qualitative, nous avons bien eu à l'esprit la conception standard du prototype selon Kleiber (1988) qui est la suivante : « le prototype est un objet singulier et une image mentale élaborée à partir d'un individu qui est considéré comme le meilleur exemple d'une catégorie dans la conscience d'une société, et qui sert de mesure, d'étalon aux individus qui en semblent proches. Les individus proches sont prototypiques, les individus éloignés sont périphériques »⁷ (Rey-Debove 1989, p. 145). La sélection quantitative a été orientée par le domaine discursif dans lequel le NdM affiche davantage d'occurrences ainsi qu'une saillance attestée ; des domaines de prédilection de circulation des collocations prototypiques dont la base est « Canada dry » ont été définis et fixés. Nous proposons une sélection de contextes situationnels disposés en fonction de leur saillance quantitative de circulation

⁷ Pour une analyse des catégorisations périphériques des collocations prototypiques dont le collocatif est représenté par un NdM, nous renvoyons à Tonti (2023).

qui confirment que le prototype en tant qu'entité construite d'attributs typiques permet de réaliser une catégorisation sémantique assez stable et permanente au sein de sphères discursives données.

3.2. Le corpus « Canada dry »

3.2.1. Apposition du nom précédent (184 occurrences)⁸

- (1) Implacable logique que celle de la manipulation de masse, de ceux qui aujourd'hui attendent tapis dans l'ombre, pour lancer une politique xénophobe et raciste sous couvert de nationalisme canada dry, où un corse ne peut être que chrétien, français et européen. (Source : unita-naziunale.org ; datation : 5 janvier 2005 ; domaine concerné : politique).
- (2) Même s'il devait y avoir une déception électorale et en imaginant qu'elle organise une élection interne elle serait réélue avec une encore plus confortable majorité. Le seul risque qui existe c'est pour la France et pour les Français si les manœuvres de ce candidat canada dry réussissaient. (Source : francesoir.fr ; datation : non disponible ; domaine concerné : politique).
- (3) On peut toujours déplorer la récupération de la laïcité par le FN, ou "l'extrême droite", mais depuis un certain temps, tout le monde se fout de la laïcité, et cherche à la remettre en cause, et réclame une laïcité "canada dry" (cela a la couleur de l'alcool, le goût...). (Source : lefigaro.fr ; datation : non disponible ; domaine concerné : politique).
- (4) Ce qui fait monter Mélenchon, c'est la faiblesse de Hollande et de sa gauche "canada dry". (Source : nouvelobs.com ; datation : non disponible ; domaine concerné : politique).
- (5) Donc le but du jeu est de faire un truc qui ressemble, qui ait le goût, l'odeur, la couleur... Des élections canada dry quoi. Si l'imposition par la force de la démocratie avait été le but de qui que ce soit dans cette histoire. (Source : lefigaro.fr ; datation : non disponible ; domaine concerné : politique).
- (6) Ce qui explique pourquoi lesdits dirigeants poussent, comme les fonds de pension, à la maximisation de la capitalisation boursière ou au rachat des actions par l'entreprise. Même Balladur s'en est ému et a proposé une loi canada dry pour limiter les dégâts. (Source : divergences.be ; datation : non disponible ; domaine concerné : politique).
- (7) Mais voilà, la CGIL n'a pas fait cela, elle a fait l'inverse, vidant de tout son sens l'héritage de Gramsci pour n'en garder qu'une version canada dry consistant à se bureaucratiser, aller dans les institutions et l'appareil d'état. Effectivement, et comme en France, la solution anti-bureaucratique et anticapitaliste, permettant d'unir sur une base ferme une fraction majoritaire des travailleurs, de débloquent le chemin vers le pouvoir des travailleurs est bien le développement de conseils de travailleurs, ce qu'on appelle de ce côté-

⁸ Nous précisons que tous les exemples rapportés et numérotés de 1 à 51 comportant le NdM *Canada dry* ainsi que ceux qui portent sur *Port Salut* et *Chocopic* sont des énoncés authentiques circulant dans la blogosphère. Toutes les coquilles relevant de l'orthographe et/ou de la morphosyntaxe ont été retranscrites sans qu'aucune correction ne soit apportée par l'auteur de l'article.

- ci des Alpes, les coordinations. (Source : bellaciao.org ; datation : non disponible ; domaine concerné : syndicalisme, politique).
- (8) Bertrand Delanöé, lors de son élection, était hostile à la création d'une Police Municipale à Paris. Il a donc créé les ASP, sorte de Police Municipale "canada dry"⁹, essentiellement tournée vers la circulation et dont le rôle en matière de sécurité publique est donc marginal. (Source : over-blog.org ; datation : 25 mai 2013 ; domaine concerné : politique).
- (9) Turquie : je ne peux pas être objectif (pas plus que vous...) puisque je suis amoureux d'Istanbul, la plus belle ville du monde. Là je vous concède que d'années en années j'observe un plus grand nombre de femmes voilées dans les rues. Récemment le gouvernement islamiste "canada dry" a voulu imposer des mesures restreignant la consommation d'alcool. (Source : lefigaro.fr ; datation : non disponible ; domaine concerné : politique).
- (10) Alors que la France se montre, à travers les alternances, incapable d'affirmer une simple priorité au logiciel libre, dilapide notre argent dans des éléphants blancs comme le "cloud souverain", ne pratique qu'un open data "canada dry" (et ne parlons même pas de la transparence des algorithmes) ou, bien pire encore, approuve la vente de systèmes d'écoute à l'échelle d'un pays à des régimes autoritaires, comme le rappelaient encore récemment reflets.info et Libération, l'urgence semble plutôt de reconfigurer cet appareil d'État manifestement défaillant, voire prédateur. (Source : penet.org ; datation : non disponible ; domaine concerné : politique).
- (11) Quant aux électeurs bourgeois de la droite molle, ce dont ils rêvent en fait, c'est d'un FN canada dry, sans saveur et sans odeur. (Source : lefigaro.fr ; datation : non disponible : 8 avril 2013 ; domaine concerné : politique).
- (12) Pourtant, il n'est pas parvenu ni directement ni à travers Paris ou Doha à imposer un pouvoir "canada dry". La transition de la dictature vers des élections libres s'est faite sous la vigilance des réseaux sociaux qui ont empêché les représentants des intérêts tuniso-parisiens de s'installer au pouvoir. (Source : palestine-solidarite.org ; datation : non disponible ; domaine concerné : politique).
- (13) Ce brouet de politicards tricards conviendra à ceux qui ont l'habitude, voire le goût d'avaler n'importe quoi. Mais il n'y a là que glnf bis, maçonnerie canada dry... (Source : over-blog.fr ; datation : 9 septembre 2011 : domaine concerné : politique).
- (14) Cela manque d'un ou deux coups de karcher ce genre d'affirmation... et à la vôtre pour le nouvel an, avec un champagne canada dry à la façon de la relaxe de votre protégé ! (Source : over-blog.fr ; datation : impossible ; domaine concerné : société, justice).
- (15) Sur mon site mixte les premiers arrivés sont les ex assedic qui trouvent tous un poste.... les ex anpe se battent, commencent la pst 30 minutes après l'ouverture officielle.... on ne leur dit rien ! bien oui ils sont fonctionnaires canada dry, badger sur chronos c'est rapide. (Source : over-blog.fr ; datation : 27 octobre 2009 ; domaine concerné : société, monde du travail).
- (16) Beaucoup de cadres ayant évolué toute leur carrière dans des grands groupes s'imaginent que leurs fonctions de direction pourront être exercées de la même

⁹ Exemple de collocation prototypique périphérique en vertu de la présence du modalisateur « sorte de » (Tonti 2023).

- manière, ou tout au moins de manière très similaire, quand ils envisagent la reprise d'une PME. Pourtant, c'est un monde assez différent, pour ne pas dire parfois aux antipodes. Les effets "canada dry" y sont nombreux et sont autant de pièges dans lesquels il serait dommage de tomber. Annoncé comme cela, c'est une évidence. Source : lexpress.fr ; datation : 12 décembre 2012 ; domaine concerné : société, monde du travail).
- (17) C'est un accord "canada dry" qui a l'apparence d'un accord collectif découlant d'une convention de Branche mais qui, sur le fond, ne possède aucune de ses propriétés. (Source : fecfo-org-agricoles.fr ; datation : non disponible ; domaine concerné : société, monde du travail).
- (18) Une fois de plus et jusqu'à la caricature, une concertation "canada dry" qui tourne le dos aux attentes et demandes des habitants de la Métropole. La vie sociale et citoyenne est fort compromise parce que de vraies discussions édifiantes ne trouvent plus leur place dans les commissions de travail citoyennes. (Source : over-blog.fr ; datation : 11 septembre 2011 ; domaine concerné : société, vie sociale et citoyenne).
- (19) Le pouvoir d'achat est plutôt assez élevé et la "racaille" n'a pas encore gangrené le coin contrairement à d'autres villes de banlieue (ne pas se fier quoique en disent certains aux quelques casquettes des "racailles canada dry" qui restent une minorité). (Source : bulle-immobiliere.org ; datation : 5 décembre 2011 ; domaine concerné : société, vie sociale et citoyenne).
- (20) Le 3 mars 2014, le CNB a semblé revenir sur sa position. Il propose désormais de créer une blogosphère canada dry "sur une plateforme unique placée sous le contrôle de la profession" alors qu'il est si simple de maintenir la blogosphère avocats.fr des avocats sans perte de référencement, ni coût supplémentaire. (Source : blogavocat.fr ; datation : non disponible ; domaine concerné : réseaux de communication).
- (21) Gare aux études canada dry sur le Net, sur [Actu24.be/lAvenir/Le Courier](http://Actu24.be/lAvenir/Le_Courrier) [...] L'idée de ce site est venue à son créateur français Jean-Noël Lafargue à la lecture continue de vulgarisation scientifique dans les médias. (Source : savants.fr ; datation : 23 août 2007 ; domaine concerné : communication et vulgarisation scientifique).
- (22) Jeep canada dry : châssis, moteur et fabrication Fiat. Pourquoi pas après tout. Mais les tarifs sont vraiment trop salés. Le frère siamois Fiat 500x me paraît bien plus adapté à sa clientèle urbaine et féminine. Une voiture aux faux airs de véhicule de sport, tout terrain et aventureux. (Source : leblogauto.com ; datation : 23 décembre 2014 ; domaine concerné : véhicules).
- (23) Ce qui m'énerve le + c'est, comme dit un ami, tous ces véhicules "canada dry" c'est-à-dire ces soi-disant SUV. (Source : planete-honda.com ; datation : 27 avril 2008 ; domaine concerné : véhicules).
- (24) C'est oublier le succès du golden retriever, chien doux, ultra-pacifique, qui ne ferait pas de mal à une mouche... (et con comme un balai, mais ça c'est une autre histoire...). Un chien canada dry, en quelque sorte. (Source : 24heures.ch ; datation non disponible ; domaine concerné : trait de caractère).

Les collocations qui insistent sur des discours visant la politique française ou étrangère se taillent la part du lion. 70% des occurrences que nous avons repérées pour les collocations dont le collocatif « Canada dry » a la fonction d'apposition du nom précédent portent sur des discours de l'actualité

politique française et internationale remontant à l'intervalle de temps 2005-2018.

Canada dry dans sa fonction d'apposition apporte des informations complémentaires sur la base qui est tantôt constituée de sigles de partis politiques comme FN, à savoir le Front National (11)¹⁰, tantôt d'un nom désignant l'orientation politique d'un parti : « gauche » (4). Le collocatif « Canada dry » qualifie également des événements politiques comme les élections (5), la candidature d'un potentiel élu (2), la promulgation d'une loi (6), un mouvement comme le nationalisme corse (1), la franc-maçonnerie à l'obédience indétronable (13), l'action des syndicats et de leurs membres (7), le principe de la laïcité de l'État qui est récupéré de manière apparemment intéressée par le FN (3), le corps de la Police Municipale parisienne (8) ou un autre appareil d'État : le « open data de l'État » qui aux dires de l'auteur de l'énoncé serait un instrument prédateur d'information (10). Les collocations « gouvernement islamiste Canada dry » (9) et « pouvoir “canada dry” » (12) closent la liste. Dans l'énoncé (9), Canada Dry est utilisé dans un contexte de politique de restriction de la consommation d'alcool. Du coup, le « gouvernement islamique ‘Canada dry’ » est un gouvernement qui pratique la politique de prohibition de l'alcool comme les Etats-Unis dans les années 1920.¹¹

Un autre macro-domaine où la collocation comportant le collocatif « Canada dry » a une saillance quantitative importante (20 %, 36 occurrences sur 128 au total pour cette fonction) est celui de la société qui se décline dans le sous-domaine de la justice, du monde du travail et de la vie sociale et citoyenne dont nous regardons de plus près les cotextes. Si des « fonctionnaires “canada dry” » se démarquent pour leur fainéantise (15), l'application des fonctions de direction des grands groupes à une PME aurait des « effets “canada dry” » (16) tout aussi inefficaces qu'un accord de branche entre les représentants des entreprises et une ou plusieurs organisations syndicales représentatives au sein d'une branche professionnelle (17). L'inefficacité des concertations des commissions citoyennes à l'égard des attentes des habitants de la Métropole est également manifeste selon l'énoncé (18) car tout comme le « Canada dry » il n'y a qu'une apparence illusoire de discussion édifiante qui subsiste. Ces faux airs se reproduisent dans l'énoncé (19) où des casquettes synonymes de « racaille » dissimuleraient un train de qualité de vie et de pouvoir d'achat plutôt élevés (19). Le dernier énoncé de cette batterie d'exemples relevant du domaine de la société dans lequel « Canada dry » intervient porte sur la

¹⁰ Nous précisons désormais uniquement le numéro entre parenthèse de l'énoncé présenté dans l'analyse.

¹¹ Dans ce contexte, l'expression Canada Dry renvoie directement à la publicité des années 1970. <https://www.youtube.com/watch?v=iLcEvJP9f-k>.

justice (14). Le collocatif « Canada dry » sémantise la base qui correspond à un autre nom de boisson (du champagne) en l'appauvrissant qualitativement. Cela serait à l'image de la relaxe du protégé dont il est question, donc d'une décision prise par une juridiction pénale de jugement, qui déclare non coupable le prévenu traduit devant elle. Cependant, il s'agirait d'une décision fautive et trompeuse selon l'auteur de *over-blog.fr* à l'instar du « champagne "canada dry" » qui a perdu son authenticité. L'espace par définition de circulation des ressources juridiques dont il est question dans l'énoncé (20) est celui de la blogosphère qui semble être mise à mal pour son manque de sérieux à cause du changement de plateforme annoncé dans l'exemple. La collocation « blogosphère "canada dry" » (20) a trait au domaine de la communication par réseau et à la vulgarisation scientifique¹² en ligne dont on déplore l'inexistence d'une démarche scientifique, dénominateur commun à l'exemple (21). Ceci n'est pas dépourvu d'intérêt parce qu'on peut déduire une perte de sérieux, à propos de l'information et des études scientifiques conduites en Belgique : le renvoi à l'hebdomadaire belge *L'Avenir* qui qualifie les informations d'un site de vulgarisation scientifique comme des « études "canada dry" » révèle que l'expression est utilisée en France et en Belgique avec le même sens dépassant les frontières de la France. « Jeep canada dry » (22) et « véhicules "canada dry" » (23) sont les seules attestations du collocatif « Canada dry » en apposition à un NdM qui désigne une voiture de sport, tout terrain et volumineuse comme un véhicule SUV (23) sans en être un. Pour conclure ce volet, les fausses apparences de la boisson à l'origine du NdM sont transférées à une race canine (24) dont le trait de caractère particulièrement doux dissimulerait une certaine bêtise, aux dires de l'auteur de l'énoncé.

3.2.2. *Attribut du sujet : 113 occurrences*

- (25) Des réponses unitaires sont absolument indispensables, mais une alliance politique avec le réformisme sur le pire des terrains pour envoyer des nomenclaturistes dirigeant dans les institutions électives de l'appareil d'état, c'est le canada dry de l'unité. (Source : non disponible ; datation : non disponible ; domaine concerné : politique).
- (26) Le CNSR, où sont majoritairement implantés des représentants de l'administration, des forces de l'ordre et des associations de victimes de la route, c'est le canada dry de la concertation. Ça ressemble à de la concertation, ça a le goût de la concertation, mais ça n'en est pas ! (Source : blogs.motomag.com ; datation : 30 mars 2017 ; domaine concerné : politique).
- (27) Parce que le FN (Le Pen) est le canada dry de la politique. Il se fait passer pour avoir le goût du (ou des valeurs du) christianisme mais c'est tromperie sur la

¹² Nous avons attesté 9 occurrences au total du collocatif « Canada dry » en fonction d'apposition de noms qui relèvent du domaine de la communication et de la vulgarisation scientifique en ligne ou dans les médias.

- marchandise... “ (Source : cite-catholique.org ; datation non disponible ; domaine concerné : politique).
- (28) Des méchants et des envieux disent que notre régime politique est un canada dry de démocratie. Nous leur clouons le bec en entonnant à gorge déployée notre hymne favori, repris en chœur par nos innombrables amis dans le monde entier, selon lequel nous sommes la “seule démocratie du Moyen Orient”. (Source : newsnet.fr ; datation : 14 mars 2012 ; domaine concerné : politique).
- (29) Attention, le fait de téléphoner, de relancer par mail, d’envoyer une note et autres activités commerciales de suivi sont parfois d’excellents moyens de se bercer d’illusions : c’est le canada dry de l’activité commerciale. (Source : legrandblogdelavente.halifax.fr ; datation : non disponible ; société, monde du travail).
- (30) Ah ça, contre les admins réseaux psychorigides... Enfin, il reste toujours la solution du ssh sur le port 443, après c’est juste de la tuyauterie. Mais je suis d’accord, ssh c’est le canada dry de l’internet... : Ben, moi c’est vraiment iTunes... que j’utilise pour accéder aux radios ou écouter ma musique. (Source : linuxfr.org ; datation : non disponible ; domaine concerné : informatique).
- (31) Ils sont le canada dry de la musique : on dirait un groupe à carrière, mais on est devant un phénomène éphémère. (Source : newwavehooker.com ; datation : 4 avril 2009 ; domaine concerné : industrie de la musique).
- (32) Bref l’émission Alien theory c’est le canada dry de la démarche scientifique : ça en a le goût, l’apparence mais A.T contient 0 % de démarche scientifique donc 0 % d’argument. (Source : nioutaik.fr ; datation : non disponible ; domaine concerné : communication et vulgarisation scientifique).

Le but de la sémantique du prototype est de décrire des zones de savoir prototypique partagé (Langacker 1987, p. 62). La représentation sémantique que la mémoire des auteurs des énoncés précités se fait de « Canada dry » en tant qu’attribut du sujet semble relativement stable car rattachée à un plus fort pourcentage d’occurrences aux discours portant sur la politique (90 % sur 113 occurrences au total pour cette construction). Le premier énoncé de cette batterie d’exemples insiste sur un problème d’unité politique : des réponses unitaires défailtantes compromettraient les institutions électives de l’appareil de l’État (25). Le parti politique Front Nationale déjà mentionné dans l’énoncé (11) en lien à une droite molle revient dans l’exemple (27) pour rappeler son attachement à des valeurs chrétiennes apparemment fausses. Des concertations inefficaces étaient tantôt déplorées dans l’énoncé (18) à l’égard des attentes des habitants de la Métropole que dans l’énoncé (26) dans le cadre d’une institution qui devrait assurer la sécurité nationale (CNSR)¹³. Quittant les discours portant sur la politique et l’administration, d’autres macro-thèmes ont été déjà évoqués comme celui de la société¹⁴ et notamment celui du monde du travail (16) car des tâches inadaptées au sein d’une activité

¹³ Centre National de la Sécurité Routière.

¹⁴ 7% des occurrences au total sur 113 occurrences.

commerciale distraient du but premier qui est celui de vendre à tout prix (29). Une démarche scientifique trompeuse est encore à blâmer mais cette fois-ci à la charge d'émissions télévisées (32, déjà évoqué en 20). Le monde de la musique et des protocoles informatiques (SSH) est représenté pour la première fois ayant un caractère trompeur (respectivement en 31 et 30).

3.2.3. Complément circonstancié de comparaison : 69 occurrences

- (33) Est-ce que le gouvernement colombien a le souci de dialoguer avec ces insurgés ? Le capitalisme mondial actuel par rapport à la démocratie c'est comme le canada dry par rapport à l'alcool : il a les apparences de la démocratie mais ce n'est pas de la démocratie ! C'est une ploutocratie ! (Source : rfi.fr ; datation : non disponible ; domaine concerné : politique).
- (34) Ce blog pue la droite !!! c comme le canada dry...ça a la couleur d'une critique envers Sarkozy...mais non ! (Source : typepad.fr ; datation : 28 août 2006 ; domaine concerné : politique).
- (35) Ces possibilités sont pour l'instant des menaces qui deviennent répétitives à l'encontre de Macron... il devrait consulter la carte de son ciel. Ce que vous illustrez (très bien) s'appelle post-démocratie ou post-république c'est comme le canada dry, c'est doré comme la démocratie, ça sonne comme la démocratie, mais ce n'en est pas. (Source : rfi.fr ; datation non disponible ; domaine concerné : politique).
- (36) Les députés droite populaire, sont en fait des députaillons, souvent élus dans les régions où le FN est bien implanté, qui se servent des idées du FN, comme le canada dry s'est servi des idées de la bière. C'est du faux, de la contrefaçon, qu'il faudra taxer le moment venu. (Source : lefigaro.fr ; datation non disponible ; domaine concerné : politique).
- (37) Côté défauts : la prosodie du texte ne se base que sur les allitérations ou presque, c'est bavard, très bavard, jusqu'au point d'être lassant, d'autant qu'au niveau du message, c'est un peu comme le canada dry... On a l'impression qu'il y a un message politique, quelque chose de nouveau et de révolutionnaire, mais en fait, pas vraiment, c'est un peu creux quand même. (Source : oniris.be ; datation : 14 septembre 2012 ; domaine concerné : politique).
- (38) La démocratie US c'est comme le canada dry, ça ressemble (de loin) à la démocratie, ça a la couleur de la démocratie, mais ce n'est pas de la démocratie. (Source : typepad.fr ; datation : 20 janvier 2009 ; domaine concerné : politique).
- (39) Je me demande parfois de quelle "jeunesse" sont représentatives ces masses de gosses que je sens parfois un peu moutonnaire et conservatrice - aussi paradoxal que cela puisse paraître, c'est comme le canada dry, ils ont l'aspect et la couleur (c'est le cas de le dire) de la rébellion permanente. (Source : lefigaro.fr ; datation non disponible ; domaine concerné : société ; état d'esprit).
- (40) Vivre ensemble n'est qu'un moyen, et sans amour ce n'est pas forcément nul ou "ch...t" mais cela demande toute une organisation et plein de choses qui sont comme le canada dry - il y manque l'essentiel, il n'est dans le meilleur des cas que plagié, ce qui ramollit ou asphyxie, pollue le cœur et ne fait illusion qu'auprès de ceux qui ne savent pas ou plus ce qu'est le véritable

- amour. (Source : cite-catholique.org ; datation : 30 mai 2020 ; domaine concerné : société ; état d'esprit).
- (41) En plus, les portes d'entrées un peu équivoques, elles sont déjà en place depuis longtemps. Ça s'appelle par exemple auditeur libre à l'ENS. C'est comme le canada dry, ça a le goût de l'ENS, mais c'en n'est pas... (Source : bulle-immobiliere.org ; datation : non disponible ; domaine concerné : société).
- (42) Cédric : Tu n'as jamais été un "modérateur". C'est comme le canada dry : tu en avais la couleur, mais c'est tout. Tu ne savais rien faire des attributions d'un modérateur : déplacer les fils, renommer un fil. Même pas mettre un lien "propre". Tu disais toi même que tu n'étais qu'un animateur. Source : inconnue, datation : non disponible ; domaine concerné : société ; état d'esprit).
- (43) Ce vin semble numérisé. Fait de bits et de bracs, à classer dans la catégorie technologique. Ses concepteurs jouent sur les faux-semblants : un peu comme le canada dry, ce produit a la couleur du vin, l'odeur du vin, certains croiront même qu'il a le goût de vin, mais ce n'est pas du vin. (Source : 20minutes-blogs.fr ; datation : non disponible ; domaine concerné : industrie vitivinicole).
- (44) Airbus et Airbus Military. Ça ressemble un peu à Dacia et Renault ou comme le canada dry, ça a le nom d'Airbus mais ce n'est pas de l'Airbus. Heureusement pour tous les voyageurs qui prennent ses avions commerciaux chaque jour. (Source : lopinion.fr ; datation : 3 avril 2015 ; domaine concerné : industrie de l'aviation).
- (45) Annie m'a montré ce nouveau restaurant décoré d'une grande mosaïque, passage Sainte Catherine à Bordeaux. Oui mais je crois bien que, tout comme le canada dry : "ça ressemble à de la mosaïque, c'est coloré comme de la mosaïque, mais ce n'est pas de la mosaïque". (Source : Canalblog.com ; datation : ; 16 novembre 2015 ; domaine concerné : industrie du bâtiment).
- (46) Cela fait 50 ans et plus que l'industrie japonaise copie et colle ce que fait de manière toujours excellente (certes un peu cher) l'industrie allemande... ici avec ce fuji...le leica est copié... mais c'est comme le canada dry.... on dirait mais ce n'est pas ! (Source : nikonpassion.com ; datation : 20 septembre 2010 ; domaine concerné : industrie des appareils photos).
- (47) Mettre un écran plat me déprime au plus haut point, un peu comme le canada dry et l'alcool pour faire court. (Source : Le Figaro ; datation : 5 décembre 2018 ; domaine concerné : vie au quotidien).
- (48) Les plantes ont pris une jolie teinte pourpre et sont bien cristallisées, avec une bonne odeur fruitée que je ne saurai décrire, reste plus qu'à goûter tout ça pour se faire une idée de ce que ça vaut, en espérant que ce ne soit pas comme le canada dry. (Source : lamainverte.org ; datation : 10 mars 2013 ; domaine concerné : vie au quotidien).
- (49) Ce soir très belle musique et chanson de 1982 que j'aime beaucoup je n'ai pas trouvé la version extended mais un live qui dure plus de 7 minutes donc ça y ressemble ça le goût d'un maxi mais ce n'est pas un maxi... c'est comme le canada dry puis la version simple avec le clip. (Source : eighties.fr ; datation : non disponible ; domaine concerné : loisirs ; musique).

La politique étrangère est évoquée à nouveau dans l'énoncé 33 (déjà en 10), il s'agit notamment du gouvernement colombien qui devrait s'occuper d'assurer une véritable démocratie au lieu de faire les intérêts du capitalisme

mondial. Cette continuité thématique traverse deux contextes supplémentaires différents : la France du Président Macron (35) et les États-Unis (38). Ces vagues de démocratie instable sont suivies dans l'exemplier par un message politique également vide (37). Partis politiques (36) et anciens Présidents de la République française (34) sont qualifiés comme trompeurs, l'orientation politique : droite, droite populaire contre gauche sont les ralliements les plus répandus¹⁵. Le macro-domaine de la société (23 % des occurrences au total sur 69 occurrences) est traversé par un dénominateur commun qui est celui du creux, du vide partagé par la jeunesse qui s'affiche comme très conservatrice (39), aucun sentiment de rébellion ne la traverserait, alors qu'aucun sentiment amoureux n'animerait des « couples canada dry » (40). Une faiblesse en termes de personnalité dans la gestion des échanges avec les autres peut s'avérer « un canada dry », tout comme le profil d'auditeur libre à l'ENS qui manque d'épaisseur et de solidité (41). Le panorama de l'industrie vitivinicole (43), de l'aviation (44), du bâtiment (45), des appareils photos (46) sont autant de réalités du monde qui s'avèrent peu substantielles. Des activités du quotidien comme écouter de la musique (49), installer un écran plat ou planter des arbres fruitiers (48) peuvent également s'avérer contrariantes (47), contre-productives (48) ou décevantes (49). « Canada dry » serait finalement le raccourci métonymique d'une activité incomplète.

3.2.4. Complément d'objet indirect du verbe : 2 occurrences

(50) Mais si le cinéma «ferme», que le whisky est prohibé, il faudra bien passer au «canada dry» - quitte à le relever d'eau de Cologne... Que sera sera... Avez-vous considéré l'idée de vous tourner vers le numérique ? (Source : popenstock.ca ; datation : 8 février 2012 ; domaine concerné : industrie du cinéma).

La collocation : « passer au canada dry » peut être remplacée par la formule plus générique de : « passer au succédané » mais le NdM en tant qu'activateur lexiculturel est plus efficace.

3.2.5. Complément de l'adjectif : 1 occurrence

(51) Tu n'as pas eu de problème pour changer le firmware ? J'ai du mal à comprendre l'intérêt porté sur des canada dry de machine qui ne sont que des pales copies de l'original... (Source : silicium.org ; datation non disponible : 24 décembre 2018 ; domaine concerné : informatique).

Finalement, protocoles, programmes et matériels informatiques reviennent avec le même programme de sens en (30) et en (50), à savoir « ssh c'est le

¹⁵ Les discours autour de la politique française et étrangère occupent 50% des collocations prototypiques où le NdM a la fonction de complément circonstancié de comparaison.

canada dry de l'internet » (30) insiste sur la spécificité du protocole réseau qui permet aux administrateurs d'accéder à distance à un ordinateur bénéficiant du même principe que l'Internet qui rend possible la connexion d'entité éloignées entre elles. Il s'agit donc d'une copie de l'Internet, son succédané. L'énoncé « canada dry de machine » (51) partage avec le précédent le concept d'ersatz d'un produit informatique de plus grande valeur.

Pour résumer, la saillance qualitative (Fuertes-Olivera 2012) et quantitative (Giacomini 2010) des collocations ayant comme collocatif prototypique le NdM *Canada dry* se concentre dans deux domaines de prédilection : la politique et la société dans son corollaire de sous-domaines que nous avons nommés ainsi : - monde du travail, - vie sociale et citoyenne, - état d'esprit. Une part résiduelle est occupée par des discours portant sur la communication en réseau et la vulgarisation scientifique ainsi que sur l'industrie de plusieurs secteurs d'activité (cinéma, musique, informatique, automobile). Son importance dictionnaire réside dans le fait que le signe linguistique « Canada dry » est devenu un synonyme culturellement partagé de « bidon » dans l'imaginaire mental des locuteurs du corpus réuni. Un signe métaphorique à valeur étendue qui ajoute à son référent initial de boisson, le caractère d'une instance mensongère et trompeuse.

3.3. « C'est comme le port salut, c'est écrit dessus », « et paf ça fait des chocapic » : des phrases polysémiques à la confluence avec des phrases situationnelles

Parmi les propriétés définitives du figement selon Klein et Lamiroy (2016) il y a une propriété que les phrases situationnelles partagent avec les proverbes, à savoir que ce sont toujours des phrases nominales ou verbales c'est-à-dire des prédications syntaxiquement autonomes. Ce genre de phrases n'autorise pas d'interprétation littérale.

Dans les années 1960, des publicités promouvant la marque *Port-Salut* avaient pour slogan « Le Port-Salut, c'est écrit dessus », la phrase situationnelle fredonne le slogan initial, gage de bonne qualité, en y ajoutant le complément de comparaison parce que la phrase sémantise désormais d'autres situations du réel, elle est prototypique au sens que nous avons vu auparavant mais cette entité abstraite ne réfère plus uniquement un fromage qu'elle désignait à ses origines, elle contamine d'autres référents, c'est une image mentale pour signifier que le résultat est irréfutable, par exemple :

(52) Donc le Nist a échoué (avec 20 m\$, des centaines d'ingénieurs et des années d'étude), et comme pour le port salut c'est écrit dessus (ou plutôt dedans, dans le rapport) (Source : reopen911.info).

La reprise anaphorique du présentatif « c'est » introduit un balancement

symétrique entre les deux membres de la phrase. Finalement, il s'agit d'une fausse symétrie car on a dans le premier membre six syllabes contre quatre dans le second, ce qui donne un effet de clausule, tout en refermant définitivement le sens. Ceci renforcé par l'écho sonore en « u » : « salut », « dessus ». Le lecteur a l'impression d'une vérité proverbiale incontestable, s'adaptant à toutes les situations : c'est neutre et généralisant.

Une autre propriété des phrases situationnelles porte sur le fait qu'elles relèvent en général d'un registre familier, ce que nous ne pouvons que confirmer. De plus, d'après la littérature elles ne préconisent pas une rigidité totale mais peuvent faire l'objet d'une variation morphosyntaxique, comme il ressort de l'usage de la forme négative dans l'exemple suivant :

- (53) je me connecte à mon jeu quelque seconde sur l'écran principal des pixels de texture bug et la seconde qui suis des trames entières lézarde l'écran avec des bugs de texture (rouge criard, vert pomme etc...) le jeu se ferme et paf... non ça fait pas des chocapic mais des problèmes... (Source : pc-infopratique.com).

Dans l'expression « faire (ou ne pas faire) des chocapic », le NdM *Chocapic*, à savoir une variété de céréales chocolatées pour le petit-déjeuner, acquiert un sens étendu métaphorique et donc polysémique qui en l'occurrence est celui de « soucis, problèmes, bugs informatiques ». La forme négative dans sa version de registre familier introduit une touche d'ironie par la forme déceptive.

Cependant, cette expression à la longévité attestée commence également à s'installer en discours avec un sens étendu conventionnel qui est celui de marquer l'accomplissement ou le résultat d'une série de gestes ou de situations, ce qui à nos yeux lui fait acquérir la fonction de phrase situationnelle. L'exemple suivant semble confirmer notre intuition :

- (54) Enfin c'est aussi parce que chez l'opérateur ils ne sont tout bonnement pas capables de voir que l'ancienne locataire (qui donc n'habite plus là, non non, nous ne faisons pas encore colocation de plusieurs ménages, non mais ho !) avait changé de numéro... Et paf... non ça ne fait pas des chocapic (dommage quand même, c'est bon c'est truc là !!), mais ça coupe notre ligne, au lieu de la leur ! (Source : blogg.org ; datation : 23 mars 2008).

Nous remarquons un effet mnémotechnique car « paf » et « chocapic » fonctionnent sur une onomatopée et la finale onomatopéique du NdM. Globalement, l'idée d'accomplissement d'une action par accumulation de passages intermédiaires est davantage attestée du point de vue des occurrences, un domaine de prédilection s'impose en termes d'usage, celui de la vie au quotidien.

Finalement, si la phrase situationnelle, un slogan à ses origines à la forme affirmative saluait un geste, une action qui se produisait avec une certaine aisance, la forme antonymique marque qu'une situation donnée peut

échouer avec la même facilité. Il y a ainsi un implicite culturel partagé par les « écrivains » natifs ou experts de la langue-culture française qui va leur permettre de faire des réglages de sens en contexte.

Ces phrases situationnelles comme les collocations pourraient renseigner un dictionnaire des collocations.

4. Conclusion

Notre but était d'étudier des collocations de noms prototypiques et des phrases situationnelles qui mériteraient de figurer dans des dictionnaires de langue générale pour leur représentativité culturelle.

Pour ce faire, nous avons fait appel à la sémantique du prototype partagé qui porte sur la représentation sémantique que la mémoire des locuteurs se fait d'une entité nommée, à savoir d'un NdM. Nous avons également convoqué la théorie du contextualisme firthien pour le concept de collocation qui mise son attention sur le co-texte du mot-clé (dans notre étude un NdM) et notamment sur les liens sémantiques et syntaxiques qui se tissent. De surcroît, les traits de la saillance informationnelle et de la saillance quantitative des collocations étudiées, à savoir la saillance de certains mots familiers dans une situation donnée a fait l'objet d'une analyse attentive. Il en découle que pour le NdM *Canada dry* nous avons passé en revue 369 occurrences qui nous ont permis de répondre aux critères d'analyse précités. Le prototype en tant qu'entité construite d'attributs typiques permet de réaliser une catégorisation sémantique assez stable et permanente au sein de sphères discursives données. Le domaine discursif de prédilection est indéniablement celui de la politique nationale et étrangère, des représentants de partis clivés plutôt à droite y figurent davantage. Le monde du travail, l'associationnisme syndical et les artefacts : produits de l'art (musique, cinéma) et de l'industrie (informatique, communication audiovisuelle) correspondent à des sphères discursives quantitativement fort représentées aussi. Les liens syntaxiques que le nom prototypique « Canada dry » tisse avec ses co-occurents sont multiples. Nous avons remarqué qu'en termes de fréquence, les fonctions syntaxiques que le signe linguistique « Canada dry » occupe dans la structure d'ensemble de la phrase où il est employé sont respectivement en apposition du nom précédent et d'attribut du sujet. Du point de vue sémantique, il est indéniablement devenu un synonyme culturellement partagé d'instance mensongère et trompeuse dans l'imaginaire mental des locuteurs du corpus réuni.

« Véridique, tel le port salut c'est écrit dessus », à savoir, l'évidence en tout acquiert une réplique aux arguments indétronables. Si la stabilité sémantique en discours est assurée par le signe linguistique *Port Salut*, il n'est pas possible de souligner une saillance quantitative importante du point

de vue de la linguistique de corpus, le corpus *frTenTen17* recense à peine une vingtaine de contextes, a marque n'est plus aussi connue qu'autrefois. Néanmoins, son implicite culturel trotte encore dans l'esprit des locuteurs plus âgés.

La dernière phrase situationnelle que nous avons analysée : « et paf ! Ça fait des chocapic ! » profite d'une longévité attestée en discours qui continue d'être nourrie par les locuteurs, ce qui se traduit par une saillance quantitative solide.

Il n'en demeure pas moins que cette ébauche de réflexion mérite d'être approfondie en utilisant d'autres noms de marque comme banc d'essai pour vérifier si les notions de prototype et de collocation retenues sont des cadres solides à des fins lexicographiques. La ressource des corpus est indéniablement très convaincante pour mettre en valeur la créativité des locuteurs, la souplesse du NdM et pour valoriser des ressources qui mériteraient d'être recensées à des fins dictionnaires.

Bionote : Michela Tonti est enseignante-chercheuse auprès de Università degli Studi di Bergamo. Docteure en Science du langage auprès de Università di Bologna depuis mars 2019, elle est l'auteure d'une monographie parue en avril 2020 aux éditions de L'Harmattan. Le titre de son ouvrage est *Le nom de marque dans le discours au quotidien : prisme lexiculturel* et linguistique préfacé par John Humbley. Ses domaines de recherches sont : la linguistique de corpus, l'analyse des discours médiés par ordinateur, la sémantique référentielle, l'onomastique commerciale en perspective lexiculturelle, linguistique et juridique. La terminologie et notamment la néonymie afférant aux domaines du droit européen, de la comptabilité et de l'égalité F/H (femme/homme) ont fait l'objet de plusieurs travaux. Depuis 2021 elle s'intéresse à la traduction neuronale, elle a notamment collaboré au projet européen *Artificial Intelligence for European Integration* avec sa participation aux groupes italophones et francophones du projet *Empowering Multilingual Inclusive Communication - E-MIMIC* en collaboration avec l'Université de Bologne et l'Ecole Polytechnique de Turin. Toujours dans le cadre de ce même projet, elle collabore actuellement à l'implémentation de la version en français de France du dispositif *Inclusively*.

Adresse de l'auteure : michela.tonti@unibg.it

Références bibliographiques

- Altmanova J. 2013, *Du nom déposé au nom commun : néologie et lexicologie en discours*, EDUCatt, Milano.
- Altmanova J. 2018, *Les dérivés de l'onomastique commerciale entre brandverbing et créativité des locuteurs*, in Gérard C. et al. (éds.) 2019, *La néologie en contexte : cultures, situations, textes*, Éditions Lambert Lucas, Paris, pp. 231-244.
- Altmanova J. et Le Tallec G. 2019 (éds.), *Lexicalisation de l'onomastique commerciale. Créer, diffuser, intégrer*, Peter Lang, Bruxelles.
- Corpus frTenTen, <https://www.sketchengine.eu/frtente-french-corpus/>
- Bartsch S. 2004, *Structural and Functional Properties of Collocations in English. A Corpus Study of Lexical and Pragmatic Constraints on Lexical Cooccurrence*, Verlag Gunther Narr, Tübingen.
- Berthelot-Guiet K. 2003, « Ceci est une marque ». *Stratégies métalinguistiques dans le discours publicitaire*, in « Communication et langages » 136, pp. 58-71.
- Bothma T.J.D. and Tarp S. 2012, *Lexicography and the relevance criterion*, in “Lexikos” 22, pp. 86-108.
- Condamines A. et Narcy-Combes J.-P. 2015, *La linguistique appliquée comme science située*, in Carton F., Narcy-Combes M.-F. et Toffoli D. (éds.), *Cultures de recherche en linguistique appliquée*, Riveneuve éditions, Paris.
- Di Vito S. 2012, *L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE*, in « Linx » 68-69. URL : <http://journals.openedition.org/linx/1519> (07.12.2023).
- Fèvre-Pernet C. et Roché M. 2005, *Quel traitement lexicographique de l'onomastique commerciale ? Pour une distinction Nom de marque/Nom de produit*, in “Corela” hors-série.
- Firth J.R. 1957, *Papers in Linguistics 1934-1951*, Oxford University Press, Oxford.
- Fuertes-Olivera P.A. and Nielsen S. 2011, *The dynamics of terms in accounting: what the construction of the Accounting Dictionaries reveals about metaphorical terms in culture-bound subject fields*, in “Terminology” 17 [1], pp. 157-180.
- Fuertes-Olivera P.A. et al. 2012, *Classification in Lexicography: the concept of collocation in the Accounting Dictionaries*, in “Lexicographica” 28, pp. 293-307.
- Galisson R. 1993, *Les palimpsestes verbaux : des révélateurs culturels remarquables, mais peu remarqués*, in « Repères. Recherches en didactique du français langue maternelle » 8, pp. 41-62.
- Galisson R. et Coste D. 1976 (dir.), *Dictionnaire de didactique des langues*, Hachette, Paris.
- Giacomini L. 2010, *A proposal for an electronic dictionary of Italian collocations highlighting lexical prototypicality and the syntactic-semantic relations between collocation partners*, in Dykstra A. and Schoonheim T. (eds), *Proceedings of the 14th Euralex International Congress*, Fryske Academy, Leeuwarden, pp. 1183-1192.
- Grossmann F. et Tutin A. 2003 (éds.), *Les Collocations : analyse et traitement*, De Werelt, Amsterdam.
- Kleiber G. 1988, *Prototype, Stéréotype : un air de famille ?*, in « Documentation et recherche en linguistique allemande contemporain » 38, pp. 1-61.
- Kleiber G. 1990, *La sémantique du prototype. Catégories et sens lexical*, PUF, Paris.
- Lamiroy B. et Klein J. 2016, *Le Figement. Unité et diversité. Collocations, expressions figées, phrases situationnelles, proverbes*, in « L'Information Grammaticale » 148, pp. 15-20.

- Lamiroy B. 2016, *For a typology of phraseological expressions: how to tell an idiom from a collocation?*, in Orlandi A. and Giacomini L. (eds.), *Defining collocation for lexicographic purposes. From linguistic theory to lexicographic practice*, Peter Lang, Bern.
- Langacker R.W. 1987, *Foundations of Cognitive Grammar*, Stanford University Press, Stanford.
- Laurent B. 2010, *Nom de marque, nom de produit : sémantique du nom déposé*, L'Harmattan, Paris.
- Mel'čuk I. 2012, *Semantics: From meaning to text. In the series Studies in language companion series*, John Benjamins, Amsterdam/Philadelphia.
- Orlandi A. and Giacomini L. 2016 (eds.), *Defining collocation for lexicographic purposes*, Peter Lang, Bern.
- Orlandi A. 2016, *Monolingual collocation lexicography: State of art and new perspectives*, in Orlandi A. and Giacomini L. (eds.), *Defining collocation for lexicographic purposes*, Peter Lang Verlag, Lausanne, pp. 19-70.
- Paveau M.-A. 2017, *L'Analyse du discours numérique. Dictionnaire des formes et des pratiques*, Hermann, Paris.
- Rey-Debove J. 1989, *Prototypes et définitions*, in « Documentation et recherche en linguistique allemande contemporain » 41, pp. 143-167.
- Rosch E. 1975, *Cognitive Reference Points*, in "Cognitive Psychology" 7, pp. 532-547.
- Tognini-Bonelli E. 2001, *Corpus linguistics at work*, John Benjamins, Amsterdam.
- Tomasello M. 2019, *Becoming human: A theory of ontogeny*, Belknap Press of Harvard University Press.
- Tonti M. 2019, *La lexiculture à l'épreuve de l'analyse de corpus : le nom de marque Ladurée ou le gage de culture*, in Altmanova J. et Le Tallec G. (éds.), *Lexicalisation de l'onomastique commerciale. Créer, diffuser, intégrer*, Peter Lang, Bruxelles, pp. 383-402.
- Tonti M. 2020a, *Le nom de marque dans le discours au quotidien : prisme lexiculturel et linguistique*, L'Harmattan, Paris.
- Tonti M. 2020b, « Ripolinage verdâtre », « ripolinage médiatique » ou « ripolinage sémantique » ? Analyse outillée du nuancier sémantique d'un nom de marque dans les médias sociaux, in « Repères DoRiF » 22, <http://www.dorif.it/reperes/ripolinage-verdatre-ripolinagemediatique-ou-ripolinage-semantique-analyse-outillee-du-nuancier-semantique-dun-nom-de-marque-circulant-dans-twitte/> (07.10.2023).
- Tonti M. 2021a, *Pour une petite fabrique de noms de métiers au féminin (et au masculin) à partir de noms de marque circulant dans les blogs*, in Fagard B. et Le Tallec G. (éds.), *Entre masculin et féminin... français et langues romanes*, Presses de la Sorbonne Nouvelle, Paris, pp. 73-101.
- Tonti M. 2021b, *Pour un dialogue interculturel : évolution sémantique du Nom de Marque Kärcher dans le discours*, in "Anales de Filologia Francesa" 29, <https://revistas.um.es/analesff/article/view/483221> (07.10.2023).
- Tonti M. 2021c, *Autour de la métaphore : « Le BHV de la pensée » ou comment décrire le réel par le biais d'un Nom de Marque*, in Contini A. et al. (a cura di), *Quaderni del CIRM*, Tab edizioni, Roma, pp. 259-287.
- Tonti M. 2023, *L'approximation catégorielle à l'épreuve du Nom de Marque circulant dans la blogosphère*, in « Langages » 229, pp. 91-114.
- Trédan O. 2010, *Itinéraire d'un blogueur : entre quête de reconnaissance et visibilité limitée*, in « Itinéraires » 3 [2], pp. 83-93.
- Williams, G. 2013, *Reviews*, in "International Journal of Lexicography" 26 [1], pp. 90-94.

USOS Y FRASEOLOGÍA DE LA PALABRA *PELO* EN CASTELLANO

Un estudio de corpus hasta el siglo XVII¹

DANIELA CAPRA
UNIVERSITÀ DI GENOVA

Abstract – This paper aims to explore the lexical combinations and the semantic relationships of the Spanish word *pelo* in texts composed between the thirteenth and the beginnings of the seventeenth century. The starting point for this research is the online corpus of the Real Academia Española CORDE. It has been investigated entering the word *pelo* and searching its meanings (with the grammatical value of subject or complement) and its frequent combinations with either verbs that give a figurative sense to the expression, or prepositions and other grammatical parts of speech. The different meanings of the collocations emerging from the research, the mutual semantic relationships and the possible variation over time to which each expression is subject have been examined. Through this survey we will see the springing up of the different meanings of the word and the consolidation in proverbs, idioms or phrases of many of these combinations. Some of them have survived up to our days, while others have not. The survey also provides keys for the study of the motivation of the phraseological expressions.

Keywords: historical linguistics; phrase; *refrán*; figurative language; metaphor.

Por cierto, Sancho –dijo don Quijote-, que siempre traes tus refranes tan a pelo de lo que tratamos cuanto me dé Dios mejor ventura en lo que deseo.

(M. de Cervantes “Segunda parte del ingenioso caballero don Quijote de la Mancha”, 1615, p. 701).

¹ Este trabajo se inscribe en el proyecto *Variación y codificación fraseológica en la historia del español (siglos XIII-XVIII) (CODIFRAS)* con referencia PID2023-152770NB-I00, financiado por el Ministerio de Ciencia, Innovación y Universidades (Proyectos Generación de Conocimientos 2023) y gestionado por la Universidad de Jaén a través del Departamento de Filología Española. El proyecto forma parte del Programa Estatal para Impulsar la Investigación Científico-Técnica y su Transferencia, Subprograma Estatal de Generación de Conocimiento y está dirigido por el Dr. Francisco Pedro Pla Colomer (Universidad de Jaén) y el Dr. Santiago Vicente Llavata (Universitat de València).

1. Introducción

Los continuos desarrollos tecnológicos en el ámbito informático de las últimas décadas han favorecido la creación de corpus textuales de diferentes tamaños, utilizables para el estudio de aspectos específicos de la lengua. En lo que atañe al español, son muy numerosos y variados los proyectos de investigación que se valen de conjuntos de documentos almacenados electrónicamente (tanto creados ex profeso, como previamente disponibles en la Red) que los estudiosos aprovechan para indagar las más diversas cuestiones. Para la averiguación de aspectos relativos a la lengua del pasado, el Corpus Diacrónico del Español (CORDE), construido por la Real Academia Española, es de suma utilidad, ya que incluye textos de todas las épocas pretéritas, incluso los que datan de los primeros inicios del dialecto castellano. Si bien es verdad que en algunos casos las versiones textuales incluidas no son conformes con los criterios filológicos óptimos (Rodríguez Molina, Octavio de Toledo y Huerta 2017), por lo general se trata de transcripciones que respetan las formas lingüísticas antiguas. La variedad de tipos textuales y la posibilidad de selección de los mismos, además de la opción de elección cronológica *a quo* y *ad quem*, permiten búsquedas circunstanciadas y coherentes con el propósito de la investigación.

Las indagaciones que se pueden llevar cabo a partir del CORDE y otros corpus diacrónicos son fundamentalmente una labor filológica asentada en el estudio de la historia de la lengua en sus diferentes dimensiones y poniendo especial atención en la cronología, cuyo propósito es observar la variación lingüística en sentido amplio (Echenique Elizondo, Pla Colomer 2021), sin olvidar el proceso de formación y eventualmente de rompimiento y disolución de la mutua vinculación de elementos léxicos que en precedencia se habían combinado habitualmente y que llegan o no a constituirse en unidades fraseológicas (para cuya denominación y características nos basamos en Corpas Pastor 1996). En cuanto al fenómeno fraseológico, la presencia de dichas unidades – además de escudriñarse en los corpus – se puede encontrar registrada tanto en los repertorios paremiológicos, como en algunos diccionarios bilingües (San Vicente 2022) y en vocabularios antiguos.

En el presente trabajo, nuestro objetivo es estudiar la relevancia y los significados de la palabra *pelo*, y sus combinaciones frecuentes con diversos elementos del discurso, a partir de dicho corpus diacrónico.² Se tendrán en cuenta las diferentes tradiciones discursivas (Kabatek 2017), apuntando a lo largo de estas páginas las informaciones pertinentes. Como se verá a

² La riqueza léxico-semántica y fraseológica actual (cf., por ejemplo, la que reporta el DLE en línea) va asomando poco a poco en los textos recogidos en el corpus, sin llegar a dar cuenta de todos los procesos constitutivo de cada una de las unidades fraseológicas.

continuación, por lo que atañe a la fraseología, es temprana la formación de locuciones y corre pareja con los significados figurados; se verá también cómo, desde una locución nominal con valor referencial que se combinaba con un pequeño número de verbos, a lo largo del tiempo se pasa a su metaforización y a la simplificación de la locución misma, que queda reducida al sustantivo, originando así diferentes combinaciones verbo-nominales que perduran largamente y se van diversificando gracias al empleo de nuevos verbos. No resulta siempre posible fechar con exactitud todos estos fenómenos, ya que a su vez muchas obras – en particular las de los siglos XIII y XIV, pero no solamente estas – no tienen fecha cierta.

2. Usos rectos y usos figurados de la palabra *pelo*

2.1. Primeros testimonios

La palabra *pelo* viene del latín *pīlus*, *-i*; en esta lengua significaba tanto pelo de la cabeza, como pelo del cuerpo. Además, podía tener sentido figurado y aludir a algo de escasa o ninguna importancia. En castellano es una voz patrimonial, que aparece ya empleada en el verso 1241 del monumento de la literatura escrita en esta lengua, el *Poema de mio Cid*, donde tiene como referente un pelo de la barba del mismo protagonista, quien, para manifestar su pena y dolor, se niega a cortarse incluso un solo pelo de la barba mientras el rey Alfonso VI no reconozca su inocencia y lo vuelva a integrar a su servicio (“nin entrarié en ella tigeria ni un pelo non avrié tajado”, Montaner³ 1993). La obra es de autoría incierta y no se conoce la fecha exacta de composición, que debe situarse entre los últimos años del siglo XII (quizás el mismo 1199) y 1207, fecha que aparece en el colofón del manuscrito que lo ha transmitido.

A lo largo del siglo XIII se multiplican los casos de uso en sentido recto de la palabra, que manifiesta el abanico de significados que tenía en latín, a saber cabello de la cabeza, pelo humano y pelo de animal (incluso como colectivo), vivo o muerto.⁴ El CORDE documenta numerosos ejemplos en este siglo, con un incremento en los siguientes, como es natural, ya que aumenta la cantidad de textos presentes en el corpus. Del número total de casos allí catalogados hay que eliminar la forma del verbo *pelar* conjugada en primera persona y los dobles, o sea textos que aparecen citados dos veces,⁵

³ Se citan las obras por el apellido del editor moderno, no por el del autor del texto.

⁴ Cerca del sustantivo van apareciendo numerosos adjetivos calificativos, como por ejemplo *lezne* (hoy desusado), *suave*, *duro*, *recio*, *liso*, *lustroso*, *luciente*, *largo*, *corto*, *osco*, *erizado*; en cuanto a los colores, *blanco*, *oscuro*, *negro*, *cano*, *rubio*, *bermejo*, *roxo*, *lúteo*.

⁵ Se puede tratar, por ejemplo, del mismo proverbio presente en dos diferentes repertorios paremiológicos o bien de un mismo texto reproducido por dos editores modernos.

además de un evidente caso de errata en la anónima obra *Bocados de oro* de hacia 1250, donde se lee *pelo* (Crombach 1971, p. 123) por *yelo* (o sea, *hielo*) y las realizaciones en otros dialectos históricos, que adquieren un significado diferente.⁶ En el *Universal Vocabulario en latín y en romance* de Alfonso Fernández de Palencia – impreso en 1490 – el autor pone *pelo* (Lozano López 1992) entendiéndolo *peleo*. Por consiguiente, un análisis cualitativo como el que hemos llevado a cabo y del que damos cuenta en estas páginas reduce los datos cuantitativos ofrecidos en el CORDE.

Entre los documentos que mencionan la palabra más frecuentemente, se encuentran algunos textos como el *Libro de los animales que cazan* (una traducción del árabe llevada a cabo hacia 1250; Cárdenas s.f.) o el anónimo *Libro de los caballos* (de hacia 1275; Herrera, Sánchez 2000), dos tratados – de cetrería el primero y de veterinaria el segundo – en los cuales *pelo* se repite respectivamente 5 y 11 veces. En ambas las acepciones tienen que ver con el mundo animal, como es de esperar; otras menciones semejantes aparecen esporádicamente en otros textos. En la segunda mitad del siglo XIII se hallan tanto una precisión sobre el color del pelo de un animal vivo (“lexo a Perico de Peña, fijo mio, una mula de pelo moreno”, Sánchez-Prieto 1999),⁷ como algunas referencias al concepto colectivo de *pelo del animal muerto*, donde se oponen los “cueros con pelo” a los “cueros tanyados” o “tanados” (Sendra Cendra 1966), o sea curtidos, como se lee en la anónima *Relación de los productos del peaje de Alagón* (1291). El mismo concepto puede estar expresado brevemente como “cuero tanado o peloso”, como acontece en el *Fuero de Navarra* (fechado entre 1250 y 1300) en tres ocasiones (Sánchez-Prieto Borja 2004). Estos significados de la palabra *pelo* se encuentran documentados también en textos de las centurias siguientes y permanecen hoy en día, como es sabido y se puede leer en los diccionarios de la lengua española, como por ejemplo el DLE en línea. Más rara resulta la acepción relativa al pelo humano, presente tan solo en dos textos más de la centuria.

Un caso curioso es el del compuesto sintagmático con valor nominal *pelo malo*, que identificaba lo que ahora se conoce como plumón (de las aves), o sea, como reza el DLE en la acepción 9, “pluma fina de las aves debajo del plumaje exterior”. Lo recoge el *Libro de Alexandre*, un poema épico en alejandrinos compuesto en la primera mitad del siglo XIII, quizás en la segunda o a más tardar la tercera década del siglo (Rodríguez Molina, Octavio de Toledo y Huerta 2017, p. 13), en una graciosa descripción de la

⁶ Es el caso de la combinación de la preposición *per* y del artículo *lo* (*pelo*), presente en textos gallegos y de área leonesa que se observa en trece ocasiones.

⁷ Seguimos en esto el DLE, que distingue esta acepción de otras parecidas (como la número 2: “Conjunto de los pelos” y la 3: “Cabello de la cabeza humana”) y la fija con el núm. 5, definiéndola “Capa o color de los caballos y otros animales”. Con el mismo sentido se empleaba “pelaje”, que Covarrubias (1611) s.v. “pelage” define “la color del pelo, y el lustre”.

época primaveral: “son los paxariellos de mal pelo exidos, los távanos que muerden non son aún venidos” (Cañas 1988).⁸ En los siglos siguientes, como se verá más adelante, esta idea de la muda será aprovechada figuradamente.

2.2. Un símil y una metáfora

Si pasamos a los usos figurados, son dos las obras tempranas en las cuales se puede apreciar un matiz que va más allá de la literalidad; ambas pertenecen a la última década del siglo XIII y son anónimas, pero su composición fue impulsada por el rey Sancho IV, hijo de Alfonso X. Se trata de la *Gran conquista de Ultramar* y de los *Castigos del rey don Sancho IV*. La primera es una recopilación en prosa de fuentes historiográficas y épicas francesas y occitanas (Bautista 2005); la palabra *pelo* se repite tres veces en contextos casi idénticos: “en el conseio que el daua, aiue mesclado del pelo del lobo” (Cooper 1995, fol. 208v); “el maestre del temple dixo que aun auie y del pelo del lobo” (Cooper 1995, fol. 209r); “el maestre del temple dixo que aun auie y buelto del pelo del lobo & el conde fizo como que no lo oyera” (Cooper 1995, fol. 209r); el sentido del comentario es negativo y tiene que ver con el engaño y la disimulación. Parece probable que la expresión figurada, que se reitera en el espacio de un solo folio, sea una consecuencia de la fuente de la cual se vale el compilador; en efecto, en castellano el sintagma *del pelo del lobo* no tiene otros testimonios (por lo menos, en el CORDE) y además la presencia del partitivo *del* que lo introduce es propia de la lengua francesa y no de la castellana.

En cuanto a los *Castigos*, un texto sapiencial cuya primera versión es de 1292 (Bizzarri 2012), la palabra se encuentra combinada con el verbo *mudar* y tiene por referente el pelo del zorro; en el capítulo titulado “De la pasçiençia del Rey antigonio”, se evoca una anécdota protagonizada por Vespasiano, recordado por su sabia respuesta a una acusación formulada a través de un adagio que parangona el zorro al hombre:

Eso mesmo cuenta alli de uaspasiano al qual dixo vn caullero que la Raposa podia mudar el cabello mas non el coraçon. E el, que non podia mudar el pelo njn el coraçon. E el Respondio a tales omnes commo a estos deuemos les Riso E a nos castigo & a los malos, pena.

La misma anécdota volverá a aparecer en el *Libro de los exemplos por A.B.C.* (Keller 1961), compuesto a comienzos del s. XV por Sánchez de Vercial, donde además se encuentra la alusión al proverbio (p. 283):

⁸ Todas las citas textuales vienen del CORDE, de no ser que se señale lo contrario. En todas se respeta la grafía presentada en la edición, lo que puede ocasionar incongruencias, como en este caso, en que el texto ha sido parcialmente modernizado por el editor.

Item se dize de Vespasiano que deziendole un ombre: - La rraposa puede mudar el pelo. – E este dezia por quanto el era cobdiciosso de dinero e que en su vida non podia menguar la cobdiçia e avariçia. E con paciençia rrespondio e dixo: - Demos rriso a todas estas cossas e demos correpcion a nos e pena a los crimenes e delictos e males.

La afirmación sentenciosa de este párrafo coincide con la del anterior y se aplica al mismo personaje, Vespasiano; también su respuesta viene a decir lo mismo. El animal, un zorro, es parangonado con el hombre, aunque no se llega a explicitar la comparación en este segundo ejemplo. El texto más antiguo sugiere un símil que, si bien no está desarrollado completamente, está planteado de manera bastante clara; por el contrario, *El libro de los exemplos* presenta en la anécdota una afirmación donde aparece solo el primer término (*raposa*, ‘zorro’), dejando al personaje implicado (y a quienes escuchan la historia) la tarea de sacar las debidas conclusiones: hay aquí una metafóricidad del discurso, aunque enseguida la voz autorial se encarga de explicar las razones de semejante afirmación y sus implicaciones. El paso del símil del texto más antiguo a la metáfora del más moderno (del siglo XV) marca un hito importante en la construcción discursiva de una narración: si en *Castigos a Vespasiano* se le compara con un zorro, en el *Libro de los exemplos* el emperador se superpone a este animal, aunque se marcan, sin embargo, las diferencias, que dejan al animal en mejor lugar que al ser humano.

La fuente de esta anécdota compartida por ambas obras es latina y se halla en la biografía de Vespasiano que forma parte del *De vita Caesarum* de Suetonio. En el octavo y último libro de la obra, el autor cuenta que un hombre dirigió a Vespasiano, que acababa de llegar a ser aclamado emperador, la oración *Vulpēs pilum mutat, non mores*, porque este le había denegado la libertad que le hubiera podido otorgar gratuitamente; aludía con ella a la avaricia que lo había caracterizado y que seguía siendo un aspecto de su personalidad, a pesar de ser emperador. Esta obra de Suetonio gozó de gran interés a partir de la Edad Media en España, cuando la producción de textos didácticos y morales estaba en gran auge e intensa era la búsqueda de ejemplos de la antigüedad. La oración se convirtió en proverbio tanto en latín como luego en castellano, como veremos más adelante.

2.3. Locuciones

En el apartado anterior hemos visto cómo y con cuáles acepciones se empleó la palabra *pelo* hasta finales del siglo XIII, con una incursión en los primeros años del XV. Se trataba de unos usos con referente concreto, análogos a los del latín; incluso el dicho proverbial que conforma el núcleo de la anécdota es de proveniencia latina. El compuesto *pelo malo* tenía que ver específicamente con la muda de los pajarillos, y por otra parte animales como los zorros

también mudaban el pelo. Ahora vamos a ver cómo a estos valores semánticos y estas formulaciones lingüísticas se incorporan otros, y sobre todo otras, de muy diversas índoles, algunas de las cuales tendrán largo empleo, mientras que unas pocas (pero no menos interesantes) representan una suerte de hápax legómenon. Veamos primero, brevemente, estos últimos casos.

Uno de ellos es una locución verbal formulada como maldición en un contexto en el cual un personaje malvado reacciona a una pregunta de un hermitaño con un acto de habla sumamente descortés: “Don monge, mucho sodes preguntador. Mas, que avedes vos y de ver? Dios vos confonda el pelo” (Baird 1976, p. 84). La obra que la contiene, titulada *Cuento muy fermoso de Otas de Roma* y fechada entre 1300 y 1325, es una extensa narración tradicional, con fuentes tanto orientales como occidentales (Pereira 2007). Sería interesante cotejar este pasaje con el de la fuente que el anónimo autor del texto castellano estaba manejando, que podría dar pistas sobre la labor del traductor y la formulación de partida.

Pocos años más tarde, entre 1337 y 1348, don Juan Manuel, en su *Libro de la caza*, emplea la locución adverbial *a redro pelo*, equivalente de la actual *a contrapelo*. Esta locución no aparece en otros autores, ni en los diccionarios más antiguos: Nebrija, por ejemplo, no recoge en su *Diccionario* (1495, 1516) esta expresión, sino “apospelo”, reformulada como “pelo arriba” (esta última presente en otras obras contemporáneas, como la *Suma de la flor de cirugía* de Fernando de Córdoba) y traducida al latín como “*pilus aduersis*”, además del antónimo “*pilus secundus*”, que traduce como “pelo aiuso” (y “pelo ayuso” en 1516). Habrá que esperar a Covarrubias (1611) para ver lematizado “redropelo” (con su variante sinonímica “rodopelo”, no registrada por otros). Parece, pues, bastante probable que en la lengua oral esta locución se emplease comúnmente, mientras que en la escritura no formaba parte de las expresiones habituales.

El caso opuesto es la locución *a pelo*, que ya registra Nebrija en el diccionario de 1495, pero que no aparece en los textos del CORDE hasta mucho más tarde. En efecto, la primera documentación de una palabra o más en general su manifestación textual es casi siempre sucesiva (a veces de muchos años) a su empleo en el habla.⁹ Los tipos de discursos que se desarrollan en la escritura son necesariamente diferentes de los orales, lo cual explica este fenómeno; sin entrar ahora en semejante intrincada cuestión,

⁹ No es raro encontrar obras donde aparecen lexemas o expresiones que antedatan las fechas que documenta el CORDE. Por poner tan solo dos ejemplos, la combinación más antigua del verbo *dar* con la locución *en vacío* no es de los años 1560-1578 como se indica en dicho corpus, sino de 1551, y lo mismo acontece con la locución *viejo de Susana*, localizada en autores como Góngora y Pineda, pero en realidad presente ya en una traducción impresa en 1551 (Capra 2015).

baste con llamar la atención sobre un tema que no se aplica solo a la presente investigación, sino que atañe en general a la historia de la lengua.

Otra locución, con función adverbial, que emerge alrededor del último cuarto del siglo XIV (y confirma lo que acabamos de apuntar), es *pelo a pelo*, cuyo sentido es literal. La utiliza Fernández de Heredia en diferentes obras, en particular combinándola con el verbo *rancar*, o sea *arrancar*. A pesar de su no abundante empleo en textos escritos antiguos, el significado denotativo, composicional, debió de ser corriente, y encuentra su realización también en un refrán – citado, por ejemplo, por Pedro Vallés en su *Libro de refranes* (1549) con el número 857: “Como te haziste caluo: pelo a pelo pelando” (Cantera Ortiz de Urbina, Sevilla Muñoz 2003).¹⁰

Una nueva acepción de la palabra *pelo* – precedida del artículo indeterminado y número singular (y por tanto con fijación sintagmática locucional) – se observa en esta anónima cuarteta de tono proverbial fechada cerca de 1350, donde se usa en el sentido de “un poco”, o “muy poco”: “la alma grande vjene/ a perder se con çelo/ quando vee que mas tyene/ su vesjno un pelo” (Temprano 1995; copla 178). Esta acepción tendrá largo uso en la lengua y se conserva hasta nuestros días; está relacionada con la acepción figurada latina, donde la palabra indicaba algo insignificante. También en los versos siguientes, que con su rima y ritmo se presentan con la estructura típica del dicho o proverbio, aparece la misma figuratividad: “De furtar por ningún arte/ pelo en ty non falle parte” (Del Piero 1971, p. 48). Dicha obra, que dispone en octosílabos los diez mandamientos y otras enseñanzas moralizantes, se titula *Doctrina de la descripción o Tractado de la doctrina*, fue escrita por Pedro de Veragüe hacia 1350, y resulta ser, junto con la anterior (sin que se pueda determinar cuál es la primera), la muestra más antigua del valor figurado atribuido al término que nos ocupa. Aquí la sintaxis poética conlleva el abandono del artículo delante del sustantivo, pero su valor sigue siendo cierto.

A propósito de sintaxis, en efecto, la locución se puede encontrar precedida de la preposición requerida por el verbo, como pasa, por ejemplo, en el siguiente pasaje, que ilustra con unos versos de un poema del *Cancionero* de fray Íñigo de Mendoza, compuesto alrededor de 1507, un caso bastante común: “que la Gloria de este suelo/ en un pelo/ no debe ser estimada” (Rodríguez Puértolas 1968, p. 296).

Otra obra en verso – esta vez de estilo y género cancioneril – presenta la misma locución; es un poema de Alfonso Álvarez de Villasandino (1340/50-1424): “Amigo señor, non vos encubrades/ d’este negoçio solamente un pelo” (Dutton, González Cuenca 1993, p. 154); el significado figurado de “un poco” (“no te hagas responsable de encubrir este asunto ni

¹⁰ El mismo Juan de Valdés veía en los refranes una valiosa manifestación de la lengua antigua.

siquiera un poco”) parece bastante claro. En el *Cancionero de Baena* (1406-1435) que lo recoge, hay además un poema del mismo Juan Alfonso de Baena donde se nota una reincidencia en el sentido figurado: “Ca non vale un pelo/ su dicho moçuelo/ más frío qu’el yelo” (Dutton, González Cuenca 1993, p. 669). Su significado aquí más bien equivale a “nada”. El uso de esta locución con uno u otro sentido se observa asimismo en obras de los siglos siguientes. También en la actualidad la locución “un pelo” puede significar ya “muy poco”, ya “absolutamente nada” (cf. el DLE).

Como se puede notar, esta locución se manifiesta primero en textos poéticos, empezando por los de carácter didáctico-moral. Su empleo en textos en prosa es más tardío; sin embargo, se encauza dentro del mismo tipo de discurso con intenciones moralizadoras. En efecto, es Martínez de Toledo quien, en su *Arcipreste de Talavera* – o *Corbacho* – de 1438, inserta la locución figurada (cuyo matiz se acerca más al sentido latino) en una oración: “pero prueba de sacar dél un pelo, e verás que es lo que te digo, salvo sy eres del partido” (Ciceri 1990). En esta obra, cuyo lenguaje tiene en muchos pasajes un sabor popular o popularizante, encontramos también el refrán “¿Cómo te feziste calvo? Pelo a pelillo el pelo levando”.

2.4. Primeros refranes

Antes de la gran estación de las amplias colecciones de paremias impresas, la circulación de boca en boca de refranes y otros tipos de dichos populares tuvo que ser enorme. Lo demuestran los numerosos manuscritos encontrados, muchos de ellos aún sin editar: Combet (1971) contabilizó treinta y siete. Algunas de estas colecciones recogían muchos refranes, como el *Romancea Proverbiorum* (Combet 1971, pp. 111-115) de mediados del siglo XIV, que contaba unos 150, o los *Refranes glosados*, que son 279 (impresos luego en Burgos en 1509 en la estela del éxito de los *Adagia* de Erasmo), o los *Refranes que dicen las viejas tras el fuego* de Íñigo López de Mendoza, marqués de Santillana – más de 700, recogidos, según el CORDE, alrededor de 1454 –, o el *Seniloquium* (Cantera Ortiz de Urbina, Sevilla Muñoz 2002), con sus 494 refranes, fechado entre 1478 y 1490 (Cantalapiedra Erostarbe, Moreno Uclés 2004).

El refrán citado en el *Corbacho* se encuentra también (con el n. 151) en la colección de Santillana, posterior de unos pocos años, y en el *Seniloquium*, con la pequeña variante de la ausencia de diminutivo en la palabra *pelo*; ambas obras comparten otro refrán, “Duelo ageno de pelo cuelga” (n. 217 en Santillana), pero en el *Seniloquium* se trata de “*mal* ageno”: esta versión a su vez es compartida por otra obra que – sin ser una colección de paremias – en un apartado recoge muchas, y en general inserta un buen número de ellas en el texto: el *Cancionero de Juan Fernández de Íxar*. Se trata, también en este

caso, de variantes en alternancia libre. En la *Tragicomedia de Calisto y Melibea* el refrán aparece formulado de esta misma manera.

El tercer refrán recogido por el marqués es el que va con el número 216: “Del lobo un pelo y ese de la frente”, que presenta en Vallés (1549) y Núñez (1555) la variante “Del lobete un pelete y esse del copete”. En cambio, el *Seniloquum* se acuerda de “Qual mas qual menos, toda la lana es pelo” (n. 418), y en el *Cancionero de Juan Fernández de Íxar* se cita “El lugar de la llaga no lo cubre pelo”. En esta obra satírica de crítica política, la combinación del verbo *cubrir* con el sustantivo *pelo* se emplea además en la descripción de unas metafóricas ovejas maltratadas por el pastor: “las trasquilas a engaño/ tantas veces en el año/ que nunca les cubre pelo” (Azáceta 1956).¹¹

El último refrán citado (“El lugar de la llaga no lo cubre pelo”), presente también en colecciones sucesivas, está expresado como una oración declarativa y en efecto, entendido literalmente, enuncia una verdad. Asimismo, en los siglos XV y XVI se escribieron o tradujeron algunas obras de cirugía (como el *Arte complida de cirugía. BNM Ms. 2.165* de hacia 1450, o la traducción del *Tratado de cirugía de Guido de Cauliaco. BNM I196* de 1493, o la de la *Cirugía mayor de Lanfranco*, de 1495) donde el autor o traductor se detenía largamente en explicar cómo la herida tenía que estar limpia de pelos para poder cerrarse sin causar problemas. Explicaciones parecidas se encuentran en la *Biblia*, que tuvo en la Edad Media diferentes traducciones a lenguas vernáculas (por ejemplo, la *Biblia del Escorial* de hacia 1300 o la *Biblia Escorial I-j-4. Pentateuco*). Hay, pues, como una coincidencia entre el saber popular y el culto acerca de esta cuestión, que opone los pelos a las heridas o llagas.

3. La ampliación fraseológica

El surgimiento de unidades fraseológicas es un proceso constante en la lengua. La frecuente repetición de una secuencia de dos o más signos lingüísticos a lo largo del tiempo produce su institucionalización, o sea su reconocimiento como forma fija. Dicho de otra manera, cuando la coaparición de dos o más palabras se hace muy común, los hablantes acaban reconociendo la combinación como una unidad, cuyo significado conocen y reconocen, y cuanto más frecuente es su empleo, más probable es su consolidación e institucionalización (Vicente Llavata 2011, p. 71).

Un claro ejemplo de esta situación es el compuesto sintagmático *pelo malo*, o su variante *mal pelo*. Este último es el que aparece en el *Libro de*

¹¹ Téngase en cuenta que “no cubrirle pelo a alguien” significa “no poder medrar o hacer fortuna” (DLE s.v. *pelo*), por lo cual el verso del poema es alegórico y metafórico al mismo tiempo.

Alexandre, citado previamente, cuyo referente era esa pluma fina de las aves recién nacidas; la acción descrita con pocas pinceladas en este poema medieval es la muda de los pájaros, o sea la sustitución de esa pelusa con las plumas. El sintagma con las palabras en este orden aparece en trece textos más.¹² Los que llevan el orden inverso son cuarenta y uno.¹³ Seguramente el sintagma nominal *mal pelo*, o *pelo malo*, con su referente concreto, era muy conocido, así como todo el mundo sabía que los pájaros sustituían este pelo con las plumas. Para describir esta muda se podía emplear los verbos *mudar*, *desechar*, *perder* y *quitarse*.¹⁴

El primer uso metafórico ligado a esta expresión nominal se documenta, de acuerdo con el CORDE, en la *Tragicomedia de Calisto y Melibea* (c 1499-1502), cuyo lenguaje – fundado en diálogos – supone un acercamiento al habla popular; la oración “Y esta puta vieja querría en un día por tres pasos desechar todo el pelo malo, cuanto en cinquenta años no ha podido medrar” (Lobera *et al.* 2000, p. 145) manifiesta la voluntad de Celestina de cambiar su condición y mejorar su forma de vivir. En la misma obra aparece el mismo sintagma nominal combinado con el verbo *mudar* y con idéntico significado. Son estos los dos verbos más frecuentes asociados al compuesto nominal, que generan lo que en palabras de García-Page se define “variante sinonímica” (1998), que se da cuando entre dos unidades fraseológicas que difieren por un solo elemento léxico hay sinonimia semántica. Con Penadés-Martínez (2006) podemos añadir que estas locuciones presentan variantes en alternancia libre, ya que su elección no está vinculada a razones lingüísticas ni extralingüísticas.

Este mismo verbo, *mudar*, es usado en la *Segunda Celestina* (1534), de Feliciano de Silva, en tres ocasiones: una con la sola palabra *pelo*, dos con el sintagma *pelo malo*, sin que se aprecie una diferencia semántica en el sentido atribuido a la expresión: eso parece indicar que en la época en que Silva escribe, las dos locuciones estaban totalmente fraseologizadas y se percibían como equivalentes. Compárense los ejemplos siguientes: “Voto a la casa santa, que aún tengo de mudar el pelo malo con estos amores, que mi amo es liberal y está caído en el lazo, y no ha de doler ni estimar el gasto” (Baranda 1988, p. 262) y “Querría concluir estas pláticas, que me parece gentil moça, y dar al diablo esta puerca de Palana, que, voto a tal [...], sabe tanta ruindad como yo, y con puta tan matrera mal puedo yo mudar el pelo” (p. 158). Este autor no es el primero en emplear la locución sin el adjetivo: ya lo había

¹² Los que el CORDE documenta son quince más, pero hay que eliminar dos por no ser pertinentes (en un caso se trata de un topónimo).

¹³ También en este caso hay que eliminar algunos por diferentes razones.

¹⁴ Este último es reflexivo e implica la acción directa del pájaro. Amplia documentación sobre la muda se encuentra en textos especializados de cetrería (donde también se usa “descañar de pelo malo”, como en el cap. XLVI del tratado de Zúñiga y Sotomayor, *Libro de cetrería de caza de azor* (Gutiérrez-Arrese 1953, p. 115) o de historia natural (en particular sobre América).

hecho Torres Naharro en 1517 en la *Comedia Soldadesca*: “¡Pese al cielo!/ Qu’el hombre mudara el pelo/ según el modo que avía,/ sino que este barrichelo/ no para noche ni día” (Pérez Priego 1994, I Jornada, vv. 5-9).

En cambio, el autor de *La Celestina* muestra una preferencia por la forma *pelo malo*, que distingue del concepto de *pelo*, limitado a los refranes y a una comparación, “[un hilado] delgado como el pelo de la cabeza” (Lobera *et al.* 2000, p. 117).¹⁵ En dos ocasiones cita Silva el mismo refrán citado en *La Celestina*, insistiendo en su popularidad.

Finalmente, en la *Segunda Celestina* se encuentra también la locución “medrar el pelo”, que combina la idea de fondo de la locución “mudar el pelo” – o sea “medrar” –, con su forma composicional; su autor pone en boca de Areúsa la oración “Yo he medrado con él tan poco el pelo como mi prima con el su mal passar de Crito” (Baranda 1988, p. 179). No parece una creación léxica del autor, ya que se documenta el refrán “No te medre más el pelo que a rocín de halconero”, citado tanto por Vallés (1549) como por Núñez (c. 1550, publicado póstumo en 1555). Asimismo, en la *Comedia Pródiga* de Luis de Miranda (de 1554), se encuentra la variante “medrar en el pelo” con el mismo significado: “La halda me traigo llena,/ verdad qu’es de cornados,/ no de lindos ducados,/ ellos y la buena estrena,/ como si fueran arena,/ me los hechava [sic] sin duelo,/ mas yo no medre en el pelo/ si también no le soy buena” (Pérez Priego 1993, p. 341).

Por los mismos años, el compuesto sintagmático se combina asimismo con otros verbos, entre los cuales *echar*, con uso figurado, como se lee en la *Comedia llamada Florinea* (1554), de Rodríguez Florián: “Allá te quedarás agora, que a tales empellones presto echaré yo el mal pelo” (Canet 1996, s. p.); el sentido es el de “despedir de sí” (DLE). El mismo autor, en la misma comedia, emplea también el compuesto *pelo malo*: “presto caerá el pelo malo”, y “aunque ella presto quiere, según veo, echar el pelo malo”. Otros verbos ocasionalmente combinados con *pelo malo* en sentido metafórico en los siglos XVI y XVII son: dejar, perder, salir de.¹⁶

Hay que tener en cuenta que la forma prevalente en sentido recto era “desechar el pelo malo”; es lógico que esta sea la primera que aparece

¹⁵ Los refranes citados en esta obra que contienen la palabra *pelo* son: “Mal ajeno de pelo cuelga” (Lobera *et al.* 2000, p. 240) y “Aunque muda el pelo la raposa, su natural no despoja” (p. 253).

¹⁶ En cuanto a “dejar”, escribe Juan de Pineda en sus *Diálogos familiares de la agricultura cristiana* (1589): “y por el pelo malo que dejaron con muestras de buen ejemplo, crían coletas y copetes de mayorazgo que fundan en parientes carnales” (Meseguer Fernández 1963-1965, vol. I p. 342). Pasando al verbo “perder”, San Juan Bautista de la Concepción (Juan García Gómez), en una sección de su *Un breve tratado para los hermanos donados* (c. 1606), desarrolla un símil entre los pajarillos y los hermano donados “que, primero que salgan a volar, pierdan el pelo malo que les dio el mundo” (Pujana 1999, p. 198). Es Esteban Manuel de Villegas el que utiliza el tercero de los verbos citados arriba, en estos versitos: “Aquí, pues, ya Cupido/ sale de pelo malo” (Alonso Cortés 1969, p. 194).

documentada también en sentido metafórico y que sea la más frecuente en absoluto, con seis ejemplos en el CORDE. Sin embargo, el mismo Fernando de Rojas emplea de igual modo el verbo *mudar*, y tras él otros autores. Es esta la segunda locución más frecuente, con cinco documentaciones. La tercera es con el verbo *caer*, que tiene tres casos de uso. Asimismo, es interesante apuntar que si miramos la diacronía, veremos que la elección de verbos que aparecen ocasionalmente suele pertenecer a épocas alejadas de los primeros usos (épocas en las que sin embargo también se siguen usando los verbos más frecuentes).

En cuanto a *mal pelo*, de los cuatro casos de uso figurado hay dos con el verbo *echar* – debidos a la pluma del mismo autor – y uno tanto con *mudar* como con *desechar*. Esta segunda locución, o sea la que invierte el orden de sustantivo y adjetivo, resulta menos frecuente y deja de utilizarse en textos escritos a partir de 1605. Otra cuestión que vale la pena subrayar es que las primeras muestras de uso de las locuciones están en obras celestinescas; no cabe duda de que con la primera de ellas se inaugura una nueva tradición discursiva, recogida en las continuaciones, pero también aprovechada en otras obras que de alguna manera han aprendido una lección de ella. Un ejemplo de eso es el *Diálogo de los pajes* (y no es casual que sea un diálogo: también lo es la *Celestina*); otro es la *Comedia llamada Florinea*, una obra por definición basada en el diálogo; estas son, sustancialmente, las obras que emplean la locución.¹⁷

Sin embargo, en la *Tercera parte de la tragicomedia de Celestina* (1536), de Gaspar Gómez de Toledo, no aparece *pelo malo*. Solo hay una alusión a un refrán en boca de Areúsa, que le pregunta a Elicia “¿Tú no as oýdo decir que avnque muda el pelo la raposa, su natural no despoja? (Barrick 1973, p. 171).¹⁸ Es la misma forma que recogerán Pedro Vallés y Gonzalo Correas en sus respectivas obras.

¹⁷ El otro género textual que aprovecha de manera figurada este concepto es la literatura religiosa o escrita por religiosos y religiosas, como Santa Teresa de Jesús, que en el *Libro de su vida* utiliza el símil del ave que, con su “pelo malo”, mientras no lo mude no puede volar, parangonada por la autora con el alma. La misma idea se encuentra en unos textos catequísticos de San Juan Bautista de la Concepción (entre los cuales el que hemos citado en la nota precedente), aplicada ya al alma, ya a su persona o a los hermanos donados, y en la *Instrucción de predicadores* de Francisco Terrones del Caño, donde el autor defiende una educación que, abogando por la simplicidad, renuncie a la escolástica.

¹⁸ En el llamado género celestinesco es muy frecuente encontrar primeras y a veces únicas documentaciones de expresiones fraseológicas vulgares, además de giros populares y otras manifestaciones de coloquialismos; la estructura dialógica de las obras y el tipo de mundo retratado contribuyen a favorecer estos usos que en parte abren pequeños resquicios sobre lo que sería la forma de hablar de la gente común. No cabe duda de que estas obras se inscriben en la misma tradición discursiva; esta propicia un lenguaje que se vale de los mismos estilemas y de recursos estilísticos parecidos.

3.1. Locuciones adjetivas

La formación de locuciones comienza con la fijación de secuencias sintagmáticas que se gramaticalizan o se lexicalizan (Montoro del Arco 2006), adquiriendo un significado a menudo figurado y no composicional; la estabilidad de ciertas combinaciones y su repetición pueden tener como consecuencia el que, a lo largo de los años, su sentido se generalice y llegue a ser menos concreto (Echenique Elizondo 2021), lo cual abre la puerta al surgimiento de asociaciones semánticas diferentes de la original y sin aparente relación con ella.

En el caso de la combinación sintagmática *pelo malo* o *mal pelo*, hemos visto cómo cobra sentido figurado en unión con algunos verbos de un mismo campo semántico, que constituyen lo que podríamos llamar un *haz locucional*. Otras locuciones son adjetivas. En particular, cuando la combinación sintagmática *mal pelo* se encuentra precedida por la preposición *de*, puede convertirse en un constituyente de una locución, como se lee en “[un animal] de color bermejo e de mal pelo” (Pérez de Tudela Bueso 1992), sintagma que se repite dos veces (p. 194 y p. 430) en la *Historia general y natural de las Indias* de Fernández de Oviedo (1535-1557) como parte de la descripción de un animal, y probablemente tiene un significado literal.

En cambio, en la oración “Entró el primero vn negro, vno chiquito, rubio, de mal pelo” (Crosby 1993), del *Sueño del Infierno* de Quevedo (1608), el valor parece figurado (como para decir “con mala pinta”). Este mismo autor, en una *Jácara*, da un paso más y llega a desautomatizar la locución,¹⁹ rompiendo además su unidad sintagmática: “Xeldre está en torre Bermeja:/ mal aposentado está,/ que torre de tan mal pelo/ a Judas puede guardar” (Blecua 1971). A la torre se le atribuye “mal pelo” porque se decía que Judas era pelirrojo y el nombre de la torre evoca este color relacionado con el pelo. En cierto sentido, Quevedo, aun aplicando a un objeto una propiedad humana, vuelve a presentar la locución como si tuviera sentido literal y las dos palabras que la forman tuvieran un significado independiente la una de la otra. El gran poeta áureo emplea también la locución *en mal pelo* (“le ve en mal pelo”, Blecua 1971, p. 414) en el *Poema heroico de las necedades y locuras de Orlando* (c 1626-1628); esta locución tenía un sentido eminentemente figurado, contrariamente a la anterior que tiene como punto de partida una descripción física: *estar en pelo malo*, según lematiza Correas (1627), equivale a no estar “en perfición” (Combet 1967).

¹⁹ El uso de la fraseología y particularmente su desautomatización es uno de los aspectos del conceptismo de Francisco de Quevedo comentado por la crítica que ahonda en los usos del lenguaje y los juegos con expresiones fijas que caracterizan en mayor o menor medida todas sus obras. Un panorama crítico exhaustivo puede verse en García Padrón, Batista Rodríguez 2016.

3.2. Fuentes lexicográficas

Hacia mediados del siglo XVI, dos repertorios paremiológicos, el *Libro de refranes* de Pedro Vallés y los *Refranes o proverbios en romance* de Hernán Núñez – ambos impresos en la estela de la boga de paremias, aunque son obras de propósito diferente – citan, además del conocido refrán “Avnque muda el pelo la raposa, su natural no despoja” (este segundo autor, con variantes mínimas), otro refrán parecido, con el verbo *desechar*: “Desechar el pelo malo en poco rato” – o *chico rato*, según Núñez. Más adelante, Gonzalo Correas, en su *Vocabulario de refranes y frases proverbiales* (1627), citará “Mudar el pelo malo” (Combet 1967), con la aclaración (con su peculiar ortografía) “Dízese de los ke están más medrados ke antes”.

En cuanto a la lexicografía general, la consulta del NTLLE en línea ofrece informaciones interesantes. Mientras que Antonio de Nebrija, como era de esperar, no registra la expresión *pelo malo* en su *Vocabulario de romance en latín*, sorprende que tampoco lo hagan obras lexicográficas sucesivas a las documentaciones de los textos que hemos citado, como el *Vocabulario de las dos lenguas toscana y castellana* de Cristóbal de las Casas, de 1570, o la de Richard Percival, de 1591, titulada *Bibliothecae Hispanicae pars altera. Containing a Dictionarie in Spanish, English and Latine* (donde aparece, además de “pelo abajo”, “pelo arriba”, traducido “against de haire”).

La primera obra de las que conforman dicho corpus en lematizar *pelo malo* es el *Diccionario muy copioso de la lengua española y francesa* de Jean Palet (1604): “pelomalo. Le poil folet des oyseaux”, mientras que es en el *Tesoro de las dos lenguas francesa y española* de César Oudin (1607) donde encontramos, además, “mudar el pelo malo” como intestación de lema, expresión a la cual siguen las traducciones: “tomber le poil folet [sic], changer ses meschants habits, se revestir de neuf”.²⁰ Es interesante notar que *habit* tenía referente concreto (vestido, traje) y *mechant* significaba *de baja calidad*, así que la expresión francesa se podía entender literalmente.²¹

A la zaga de los franceses, otros lexicógrafos posteriores incluyen no solo la palabra, sino la expresión fraseológica, empezando por Girolamo Vittori y su *Tesoro de las tres lenguas francesa, italiana y española* (1609), que de Oudin copia hasta las erratas, tanto que pone “poil *folet” en lugar de *follet*. De su mano añade la reformulación de la citada unidad fraseológica:

²⁰ Como apunta San Vicente (2022), el diccionario bilingüe es el lugar privilegiado para observar la lengua.

²¹ Por otra parte, Furetière, en su *Dictionnaire universel* (1690) elenca muchas acepciones de la palabra *pelo* y su fraseología, y precisa, sin embargo de lo que otros lexicógrafos más antiguos han apuntado en sus respectivas obras, que « on appelle poil follet la premiere barbe qui pousse. On le dit aussi de ce poil delié qui vient sur la peau du reste du corps ». Son acepciones que no hemos encontrado en el corpus en castellano.

“amender sa mauvaise fortune”, que confirma la metafóricidad de la misma. John Minsheu en su *Vocabularium hispanicum latinum et anglicum copiosissimum* (1617) llega a traducir “pelo malo” con “paupertas”.

La *Exposition de muchos vocablos castellanos en lengua ytaliana*, un paratexto que Alonso de Ulloa preparó para la edición de la *Tragicomedia de Calisto y Melibea* impresa en Venecia en 1553, se hace eco de la presencia, en la obra de Rojas, de la expresión, citada arriba, “pelo malo mudar” y la pone como intestación de lema, traduciéndola “megliorarsi” (Capra 2007), o sea *medrar*, en el sentido de “mejorar de fortuna” (DLE). La *Exposition* constituye un pequeño diccionario bilingüe monodireccional español-italiano que contiene en su mayoría lemas presentes en la *Tragicomedia*, con la traducción al italiano proporcionada por Ulloa, al que se debe también la selección léxica. Es la primera obra lexicográfica en registrar la expresión.

4. La vuelta a *pelo*

La riqueza fraseológica de *pelo malo* y *mal pelo* ha llevado a dejar de lado las indagaciones sobre la ampliación semántica y fraseológica que se originaron a partir de la palabra *pelo*. Ha llegado el momento de recuperar el hilo del discurso acerca de esta palabra.

4.1. Nuevas acepciones léxicas

Una acepción que solo se encuentra en textos especializados de medicina es la que en el DLE lleva el número 16, donde es definida como “enfermedad que padecen las mujeres en los pechos cuando están criando, por obstrucción de los conductos de la leche”. De ella diserta, por ejemplo, fray Vicente de Burgos en su traducción de 1494 del *Libro de Proprietatibus rerum* de Bartolomé Anglicum (siglo XIII), que siguiendo a Aristóteles explica “que si en las tetas ay muy grande abundancia de leche vienen a se endurecer demasidamente. E si en este tiempo creşcia en ellas pelo, seria señal de una gran enfermedad que es dicha pelosa” (Herrera, Sánchez 1999, fol 65V).²² Otras obras medievales de medicina, entre las cuales el anónimo *Arte complida de cirugía* (de la mitad del siglo XV) o el también anónimo *Tratado de patología* (de alrededor de 1500) tratan de la misma enfermedad padecida por la madre lactante con idéntica denominación.

²² Este texto de sabiduría enciclopédica dedica el capítulo LXV a las propiedades del pelo, fijándose tanto en el pelo animal como en el humano; el autor divide la materia en varios capítulos, al punto que el LXVI está dedicado solamente a los cabellos y sus propiedades. En este tratado la palabra *pelo* se repite en 86 ocasiones.

En un capítulo dedicado a la fiebre cuartana, afirma fray Vicente de Burgos que al cuarto día al enfermo se le “eriza el pelo”. Vale la pena apuntar a este propósito que, en el *Vocabulario eclesiástico* (1499), su autor Rodrigo Fernández de Santaella acerca del verbo latino lematizado “Inhorreo” explica “por ser frio y aspero o por temer mucho o por espeluznar se o erizarse el pelo” (Lozano 1992, fol. 95r): esta última acción es más bien el efecto visible de la sensación, lo cual muestra el paso del sentido recto al figurado de la expresión. El significado figurado ya está documentado en un poema anterior a 1499, debido a la pluma de Luis de Lucena, quien en su *Repetición de amores* – escrita cerca de 1495 – define el amor en términos de otros conceptos, en particular sacados de la medicina: “Es amor una sangria/ y un dulçor dissimulado/ que se abre con alegría/ y se cierra con cuidado;/ es de yel un bocado/ que se traga como miel;/ es como pelo herizado/ que haze estar colgado/ el brazo por causa dél” (Ornstein 1954, p. 73).

Volviendo al *Libro de Proprietatibus rerum*, y por tanto al ámbito discursivo de lo que se llamó “filosofía natural”, al explicar cuáles son los cuatro tipos de lepra y sus respectivas causas, recuerda el traductor que la tercera “es llamada raposina & alopiçia, porque assi como el raposo pierde su pelo al estio, por el calor de la sangre & del hígado, assí el enfermo desta passion pierde las sobreçejas y el otro pelo del cuerpo” (Herrera, Sánchez 1999, fol. 111v). Esta información es interesante, ya que nos da a entender hasta qué punto un animal como el zorro, además de aparecer en numerosas historias didáctico-morales cuyos protagonistas eran animales – tan populares durante la Edad Media y más allá – también era tenido en cuenta como término comparativo por su fisiología.

Pasando a cuestiones más amenas, parece oportuno señalar una curiosidad ligada a la forma de trabajar las pieles, además de la que ya vimos anteriormente, relativa a obras medievales (s. XIII), donde figuraban unos “cueros tanados y cueros con pelo”, que como se recordará son modalidades de tratamiento de la piel obtenida de algunos animales; a esta misma manera de tratar la piel, en el tercer cuarto del siglo XV se refieren los textos con “cueros cortidos o al pelo”²³ y en un texto de hacia 1493-1497 se reformula como “ningund cuero en pelambre ni al pelo” (Rubio Pardo *et al.* 1979, p. 229). Además de esta cuestión, resulta curioso, en efecto, que en la anónima *Relación de la dote de don Juan y la princesa Margarita* (de 1496, año en el cual se concertó el matrimonio del joven hijo de los Reyes Católicos con la aún más joven princesa de Habsburgo), figuran entre otros bienes materiales

²³ Por ejemplo en la siguiente norma: “en ninguna manera no se dé liçençia de oy en adelante a persona alguna para sacar corambres cortidas nin al pelo” (Millares Carlo y Artiles Rodríguez 1932, p. 116), que muestra además tanto el empleo de la palabra “corambre”, como el de la locución “en ninguna manera”. En un documento de 1478 aparece “cueros a pelo” (Carande 1968, p. 165), sin el artículo.

preciosos (perlas, diamantes, rubíes y joyas) “dos piezas de brocado de oro tirado, muy fino de pelo”, “ochenta varas de brocado de raso”, “una cama muy rica, de tres paños de brocado carmesí de pelo”, “una sobrecama de brocado rico carmesí pelo” y “dos paños de terçio pelo verde y damasco blanco” (Escudero de la Peña 1870, pp. 204-205); además, “una mula, con silla redonda, con su guarniçion e caparaçon de brocado carmesí, de pelo rico” (Escudero de la Peña 1870, p. 207). Este parcial listado de riquezas del ajuar del príncipe don Juan muestra los diferentes significados de la palabra “brocado”, que no era solo la tela de seda entretejida con oro y plata, sino también la que se hacía a partir de otros materiales, como por ejemplo el cuero – lo que algunos años más tarde se llamaría “guadamecí”, o sea, como reza el DLE, “cuero adobado y adornado con dibujos de pintura o relieve”. El brocado, pues, podía ser “de pelo” (a veces llamado simplemente “pelo”, sin preposición) o “raso”. El color que más gustaba en la época era el carmesí, probablemente por la disponibilidad del material utilizado para la tinte. También apreciaban el “tercio pelo”. En cuanto a la forma gráfica del término, era la que se solía utilizar con mayor frecuencia en la época: la separación léxica describe más claramente su composición, basada en tejer una trama y dos urdimbres. El terciopelo también podía ser “vellotado”, o sea, según el DLE, “rizo, no cortado en el telar”.

A caballo entre el siglo XV y el XVI (o más bien hacia la mitad del siglo XVI, según señalan Rodríguez Molina, Octavio de Toledo y Huerta 2017, p. 13), Alonso de Santa Cruz, en su *Crónica de los Reyes Católicos*, trata de los tejidos y otras cuestiones con ellos relacionadas, entre las cuales las leyes: “Otrosí, mandó que ningún tundidor fuse osado de tundir ningún paño sin que primero mirase si venía poblado de pelo o dañado (Mata Carriazo 1951, p. 176). El autor del *Vocabulario eclesiástico*, ya citado, Fernández de Santaella, explica bien el significado de este latinismo: “Tundo. dis. tutudi. me. cor. tunsum. Por golpear o ferir o majar o quebrantar o tondir que es sobre cortar el pelo del paño” (Lozano 1992). El interés por los textiles debió de ser alto, a juzgar por la cantidad de textos que tratan, desde diferentes puntos de vista, el argumento. Había, por ejemplo, quien hacía recomendaciones acerca de cuándo trasquilar las ovejas para obtener lana de mejor calidad, como el anónimo autor del *Informe de un mercader para el fomento de las manufacturas textiles y mejora de la calidad de los paños* (1495), que afirma: “por esperiència se ha allado que las lanas que se trasquilan por abril, se cae luego el pelo a los pannos que fassen della” (Iradriel Murgarren 1974, p. 352); aconseja, en cambio, el mes de junio. En otro texto anónimo, de 1498, hay una referencia a “un pelo [...] de la camisa” (Baranda 1995, s.p.); estas breves citas muestran otra acepción más de *pelo*, como componente mínimo de la fabricación de un paño o una camisa. También un título como *Libro de geometría, práctica y traza* (1580-1589), de Juan de Alcega, trata de cuestiones que guardan relación con los tejidos y la ropa; allí

aprendemos que “para cortar este boemio de seda, es necesario doblar la tela a lo ancho, poniendo la mitad de las baras sobre la otra mitad a pelo y labor” (Regalado López 2000, fol. 31 v). Esta locución se repite diecinueve veces en el texto, a menudo acompañada de la precisión “si / porque es damasco”, y probablemente se refiere al adorno del damasco, además de la dirección del pelo del tejido. Se trata, pues, de una combinación con sentido literal.

Para terminar con el ámbito de los tejidos, apuntamos que en un texto anónimo de 1501, la *Real cédula de los Reyes Católicos*, aparece la palabra *pelo* en referencia al mundo vegetal: “algodon filado y en pelo, cinco maravedis” (Morales García-Goyena 1906-1907, vol. II, p. 140). El material podía ser otro, por ejemplo lino, documentado en otra obra anónima de 1510.

Otra acepción todavía actual de la palabra, que el DLE define con el número 15 como “Raya o grieta por donde con facilidad saltan las piedras, el vidrio y los metales”, es la que se encuentra en una copia anónima de 1491 de las *Siete Partidas de Alfonso X. BNM 1766* (compuestas en la segunda mitad del siglo XIII), donde se lee: “asy como algund pelo o alguna señal de quebradura que era en la piedra” (Sánchez Prieto 2004, vol. V, 18R).

Finalmente, la acepción de pelo como “cuerpo extraño que se agarra a los puntos de la pluma de escribir y hace que la letra salga borrosa” que el DLE recoge con el número 11, encuentra su manifestación (casi su celebración se podría decir) en *La pícaro Justina* (1607) de López de Úbeda, cuyo comienzo gira alrededor de las palabras pluma y pelo, tanto que esta última aparece cincuenta y tres veces. La acepción está claramente expresada en los primeros versos de esta redondilla: “Cuando comenzó Justina/ a escribir su historia en suma,/ se pegó un pelo a su pluma” (Rey Hazas 1977, I p. 87). Además del vocablo, se hallan en estas páginas juegos de palabras (“Sopla Justina la tinta para quitar el pelo de la pluma. Mas antes que nos pope, quiero soplarle, aunque me llamen soplona”, Rey Hazas 1977, p. 104), refranes (“Cuando tuvieres un pelo más que él, pelo a pelo te pelas con él”, Rey Hazas 1977, p. 92), compuestos (a pospelo) y derivados (pelona, pelambre, pelado) que se reiteran a lo largo del texto, aparte de la fábula de la rana, que nació con pelo, y cómo lo barajó.

4.2. Nuevas locuciones

La emergencia de nuevas acepciones de la palabra *pelo* y las muestras textuales de su evolución en los usos, ligada a nuevas costumbres, modas, o al progreso en los conocimientos técnicos y científicos, es un índice del desarrollo de la lengua, que con el paso del tiempo cambia, se enriquece, se modifica. Lo mismo se observa con la fraseología, cuya tendencia es crecer e incrementar el número de unidades. Se trata, a veces, de la creatividad de un autor, pero en la mayoría de los casos se trata de un fenómeno más generalizado, que atañe al sistema lingüístico en su totalidad.

4.2.1. Locuciones verbales

Un caso de probable creación fraseológica es el de Íñigo López de Mendoza, ya recordado como compilador de refranes; sus intereses por la cultura, tanto la del pueblo, como la alta, incluso de raigambre italiana, le llevaron a escribir los primeros sonetos en lengua castellana, conocidos como *Sonetos al itálico modo*. Es en un verso de uno de ellos, el XXIV, de tema amoroso, donde modifica la locución “mudar el pelo” escribiendo: “Si el pelo por ventura voy trocando/ non el ánimo mío, nin se crea;/ nin puede ser, nin será fasta quando/ integralmente muerte me posea” (Gómez Moreno, Kerkhof 1988, p. 67). El poeta conocía indudablemente la anécdota donde se evoca la raposa y la sentencia que constituye su centro (o su *telos*, si se prefiere), y escoge otro verbo que no sea “mudar”, por marcar una diferencia desde el comienzo del poema, ya que su sentimiento de amor, que durará hasta la muerte, se presenta positivamente, mientras que las costumbres (o el “natural”, término presente en muchas versiones del refrán) tienen una connotación negativa. Si por un lado el poeta evoca dicha anécdota, por el otro se aleja de ella, ya que su firmeza es una virtud, no un defecto.

Otras locuciones, por el contrario, reflejan la lengua común. Es el caso de “venir a pelo”, todavía empleada hoy frecuentemente en la lengua oral, que encontramos en el *Diario del primer viaje de Colón*, de autor anónimo, redactado hacia 1492-1493: “todo es venido mucho a pelo, para que se haga este comienço” (Varela, Gil 1992, p. 181). Su significado equivale a “oportunamente”; su primera presencia en un repertorio lexicográfico se debe a Palet (1604), quien lo traduce al francés con la locución “à propos”. Tras él, por supuesto, Oudin, en 1607 (*bien à propos*).²⁴

En *Batallas y quinquagenas*, de Gonzalo Fernández de Oviedo (cuya fecha coloca el CORDE en el amplio periodo entre 1535 y 1552), encontramos otra locución verbal que ha llegado hasta nuestros días, “lucir el pelo”: “En fin, a la qual casa [una palabra ilegible]²⁵ bien el pelo le luzo. Avés de saber que el Emperador rrey don Carlos, nuestro señor, hizo su Visorrey en Nápoles al señor Marqués de Villafranca don Pedro de Toledo, [una palabra ilegible] a la Marquesa su muger doña María Osorio, e a los hijos [...]” (Avalle-Arce 1989, p. 407). Por el contenido de la oración, se entiende que el significado de la locución no tiene las connotaciones irónicas prevalentes hoy en día, sino que alude a los importantes cargos y prebendas

²⁴ Alrededor de 1570-1573, fray Diego de Estella, en polémica con las tendencias que nota a su alrededor, en su *Modo de predicar y modus concionandi* pone como ejemplo positivo a los antiguos doctores de la Iglesia “trayendo Escritura a su propósito, a pelo, y no a pospelo, y no torcida” (Sagüés Azcona 1951, p. 37); eso muestra la posibilidad combinatoria con otro verbo y la existencia del antónimo.

²⁵ El manuscrito no está bien conservado, condición que impide su lectura completa; el editor lo señala de la manera que hemos copiado con la cita.

recibidos por todos los miembros de la familia; su uso es figurado y expresa el prestigio y lucimiento de los de Villafranca. En cuanto al significado literal, el *Vocabulario eclesiástico* (1499) sub voce *Obesus* pone como traducción “grueso, luzio, redondo en carnes”, y Gabriel Alonso de Herrera, en su *Obra agricultura* (1513) afirma que el macho de la cabra tiene que tener “grande barua que tenga grande pelo largo luzio, liso y todo un color” (Capuano 1995, fol. 139r). El adjetivo “luzio” tenía entonces valor positivo y se podía aplicar al pelo.

El verbo *lucir*, por otra parte, ya se empleaba en sentido figurado, además de recto; es un ejemplo de ello la siguiente oración de Gonzalo García de Santa María que aparece en una obra de 1499: “que bulda²⁶ dela corte no salga: sin luzir en ella el blason y gloria de las armas de Aragon” (Pino Jiménez 2002, fol. 95v). Este uso del verbo es el mismo que se hace hoy en día. Además, no hay que olvidar la existencia del refrán “Al buey maldito el pelo le luzer” (o “le reluzer”), presente en repertorios del siglo XV y siguientes, citado también por Juan de Valdés en su *Diálogo de la lengua*, que muestra el origen de lo que ahora es una locución verbal.

Luis Milán, en *El cortesano* (1561) consigue ensartar dos locuciones en la misma oración: “Dixo la Reina: Canónigo, quedemos en paz, que no os faltará pelo en la ropa, y hacé que no le tengáis en la lengua” (Impr. de Aribau 1874, p. 260). Mientras que la segunda locución está en uso, la primera no tiene más documentaciones que esta.

Siguiendo con las locuciones verbales todavía existentes, vamos a pasar a una que, sin embargo, carece de la forma pronominal que tenía en textos coevos (y tiene hoy) y se presenta como “caerse el pelo” (y no “caérsele el pelo [a alguien]”), y cuyo significado, por consiguiente, no parece ser el mismo que hoy.²⁷ Vale la pena citar algunas líneas más del mero cotexto, para entender mejor el contexto general y aproximarnos así al sentido que tiene la expresión:

Porque quien de mucho mal es ducho, poco bien le empalaga, en especial que creo yo que este en toda su vida salió de cavallerizas y burdeles sino en mi casa, y agora piensa ya el don duelo que de ruin se cae a la persona el pelo, y que ayer entró rogando y que oy se ha de assentar mandando. Pues aguarde que se me caya el bocado de la boca por bova, que quanto mas se quisiere llamar a possession le haré yo que se quede del agalla (Canet 1996, escena XXV).

La mujer que habla, Marcelia (la alcahueta), muestra un claro fastidio hacia el hombre del que habla, Fulminato, un criado de Floriano, ávido, arrogante, abusón y fanfarrón, que ella acusa de querer aprovecharse de su buena

²⁶ Bula.

²⁷ También difiere de “caer el pelo (malo)”, locución comentada anteriormente.

disposición. Todo el parlamento (mucho más largo) está engastado de modismos. La expresión fraseológica en cuestión parece indicar que el hombre, por presumido y satisfecho de sí mismo, piensa que va a obtener lo que quiere porque lo quiere. Marcelia se propone que, cuanto más él intente adueñarse de lo que no es suyo, más hará ella para que se frustren sus deseos.²⁸

Otra locución empleada en esta obra (comparable por lo que atañe al lenguaje a las “celestinescas”, aunque por otros aspectos se diferencia) es “avisar el pelo”, que significa “tener cuidado”. Ahora es Fulminato el que habla, comentando una petición que le hizo Floriano: “¡Allá quedarás, diablo! ¡Y qué pensado tenía que avía yo de yr donde me mandava! A la fe, avise el pelo, que a mi salvo de honra y persona y ganancia, y aun sin offensa de Dios, serviréle porque me lo paga” (Canet 1996, escena XIX).

Aprovechando aún más la riqueza fraseológica de esta comedia, vemos en este parlamento de Felisino, el otro criado de Floriano, otra locución más: “Y el necio, pues, de mi amo que por mensaje de una alcahueta da lo que tiene [...] ¡Ay, os digo yo, que como se vea con pelo, que luego se haga delicada y ociosa!”. Parece claro que “verse con pelo” significa verse con riquezas, sustancia, bienes materiales.²⁹

En un texto de 1574 de un escritor mejicano, Fernán González de Eslava, el *Coloquio tercero a la consagración del doctor don Pedro Moya*, hablan entre ellos dos personajes alegóricos, Vanagloria y Adulación, deseando ir “a alguna fiesta donde los dos mejorásemos el pelo” (Rojas Garcidueñas 1958, I, p. 74); la misma locución se encuentra en las *Elegías de varones de Indias* del colombiano Juan de Castellanos, de 1589. Esta locución, cuyo significado genérico es “medrar”, es una variante de la más común “mudar el pelo” (o de “medrar el pelo”), que hemos visto anteriormente como heredera de “mudar el pelo malo”. En España no hay evidencias de su uso.

A lo ya dicho es preciso añadir que en la *Segunda parte de la vida del pícaro Guzmán de Alfarache*, de Mateo Luján de Saavedra, publicado en 1602, se lee “Quiso Dios Nuestro Señor que llegamos a la ciudad de Nápoles, lo que yo mucho deseaba por mudar el pelo viejo, que ya tenía orden de mi amo para que me vistiese a mi gusto y dinero todo lo que yo había menester” (Sevilla 2001, p. 152); el sentido de la locución parece referirse al cambio de

²⁸ “Quedarse del agalla” según Covarrubias (1611) es “quando uno queda asido o preso o frostrado de su pretension”. La imagen del bocado que se cae de la boca evoca la fábula del cuervo (con el queso en el pico) y el zorro adulator. También las primeras palabras de la cita corresponden a un dicho popular, interpretable literalmente.

²⁹ No es una casualidad que antes del euro se llamaran “pelas” las pesetas. Por otra parte, hay que considerar que el pelo abriga y cubre, y de uno que tiene dinero se dice que tiene “el riñón bien cubierto”. En las páginas siguientes, al tratar las locuciones nominales, volveremos sobre este asunto.

vestido y de las condiciones materiales de vida, lo cual coincide con la definición que trae el *Vocabolario italiano-spagnolo* de Lorenzo Franciosini, que en 1620 dice: “Mudar el pelo. Mutar il pelo, e si dice metaforicamente di chi ha migliorato il mangiare, o il vestire”. Además, pocas líneas más arriba el personaje-narrador de la novela dice de sí mismo que está “mal ropado”. De todo eso podemos concluir que en el siglo largo que va de la composición de la *Celestina* a la publicación de la *Segunda parte del Guzmán de Alfarache*, pues, la expresión ha ido adquiriendo un significado más preciso, restringido al mero bienestar físico y material; anteriormente, en cambio, se entendía como una mejora general de las condiciones de vida.

Por otra parte, a caballo entre el siglo XVI y el XVII encontramos otras dos locuciones que introducen nuevos verbos y crean nuevas formas de expresar conceptos quizás igual de nuevos: se trata de “hender el pelo” y “partir el pelo en el aire”, modismos que hoy se formulan con la palabra “cabello” en lugar de “pelo”, con el mismo sentido de “tener gran perspicacia o viveza en comprender las cosas” (DLE) y que vemos en las citas siguientes, la primera de las cuales muestra el sentido *productivo* (o sea el de tener viveza en el *hablar*), que también tenía la locución:

El Pinciano dixo entonces: Yo no entiendo esta algarauía. ¿De manera que la heroyca no consiente conceptos agudos?

Vgo dixo: Grandes, sí, pero agudos muy pocos. Y, si queréys saber la causa, acordaos que la épica es imitación de príncipes y señores grandes; y mirad que los príncipes y señores grandes hablan con grauedad y simplicidad alta; y mirad la gente menor quán aguda es en sus conceptos y dichos, que, assí como hienden el pelo, hienden la oreja con la agudeza dellos (Carballo Picazo 1953, vol. II, pp. 207-208).

Mientras que la locución “hender el pelo” se refiere al hablar agudo y sutil, con “hender la oreja” el personaje que habla con el Pinciano alude al escuchar y entender, o sea, al efecto que produce la agudeza. Sobra decir que en el umbral del barroco la valoración positiva del *concepto* abre el paso a las manifestaciones literarias basadas en esta idea.

El otro texto que inserta una nueva locución – parecida a la recién citada – fue escrito entre 1603 y 1607 por San Juan Bautista de la Concepción (Juan García Gómez) y se titula *Pláticas a los religiosos*. En él, la locución tiene el significado descrito por el DLE que acabamos de citar; el fragmento de la *Plática* donde aparece es el siguiente: “Acutus, agudo, que no se deje llevar de las tentaciones, etc.; que parta el pelo en el aire para no dejarse engañar del demonio” (Pujana 2002, fol. 109 v). En 1627, Gonzalo Correas, en su *Vocabulario de refranes y frases proverbiales*, explicará, con su grafía peculiar, el significado de una locución sinónima, “Korta un pelo en el aire”: “Alaba de mui sutil i delgado inxenio. Metáfora de la espada, ke korta un

kopo de lana o estopa en el aire de un rrevés” (Combet 1967, p. 706). Se trata, por tanto, de variantes libres de la misma locución.

Si pasamos brevemente a examinar las locuciones hoy olvidadas, encontramos primeramente “andar al pelo”, que el DLE marca como coloquial y desusada, definiéndola “andar a golpes”, que aparece al final de la dedicatoria del autor, Andrés de Poza, a su *De la antigua lengua, poblaciones y comarcas de las Españas* (1587): “Vuestra merced reciba el libro y al autor con aquella voluntad que desean [...] porque en cuanto a las demás partidas de estos reinos, no faltarán algunos poco devotos de esta lengua que andarán algo al pelo con el libro” (Rodríguez Herrero 1959, fol. 3 v).

También Cervantes de Salazar, en su *Crónica de la Nueva España*, ya citada, emplea la locución “no dejar pelo ni hueso”, cuyo sentido (hoy sustituido por locuciones sinónimas) es bastante fácil de intuir: “[...] pues no podemos ir sino en son y manera de fugitivos, a los cuales persigue tanto la fortuna que no dexa, como dicen, pelo ni hueso dellos” (Magallón 1971, vol. I p. 268). El DLE no registra este modismo, ni tampoco el siguiente, que aparece documentado en la *Historia de las Indias* de Bartolomé de las Casas en la siguiente oración: “Pluguiera a Dios que a los reyes hubiera costado cualquiera descubrimiento y población, en culquiera parte destas Indias, tantos dineros [...] porque otro pelo tuvieran sus reinos del que tienen” (Castañeda Delgado 1994, p. 2269).

4.2.2. Otros tipos de locuciones

Además de la creación de locuciones verbales, bastante numerosas y graciosas, como hemos visto, hubo en la misma época un incremento de otros tipos de locuciones, en particular de tipo adverbial y nominal. Se trata de formas seguramente comunes en los siglos XVI y XVII, y aún antes, sobre todo en el habla, pero estas no se habían manifestado en la escritura (de los textos recogidos en el CORDE, claro está) hasta una cierta fecha.

En cuanto a las adverbiales, encontramos por primera vez “contra pelo” en la *Obra agricultura* (1513) de Herrera: “[el fruto del olivo] si no alcançan a cogerlo a mano sacudan el oliuo con vna verdasca o caña a pelo y no contra pelo por que no dañe ni quiebren la rama: que donde las aporrean quiebran la rama” (Capuano 1995, fol. 91 v). Esta forma gráfica es rara, ya que el otro caso presente en el corpus corresponde a un cuento de Clarín. Algo más común en esa época resulta la forma actual de esta locución, que se encuentra escrita “(a) contrapelo” por primera vez en una obra anónima fechada entre 1548 y 1575, una traducción de la *Cosmografía* de Pedro Apiano (cuya versión original, en alemán, se publicó en 1524).³⁰ En lugar de la locución

³⁰ Las tres concordancias que ofrece el CORDE hasta 1600 de “a contrapelo” correspondan exactamente al mismo texto, que sin embargo se atribuye a tres obras diferentes, a saber: la

actual, con preposición, sin embargo, se suele encontrar la forma compuesta “contrapelo” con el mismo significado, como se ve en este pasaje que condensa una gran cantidad de locuciones: “poniendo, quitando y assentando la pluma y mirando a una parte y a otra, al sol, a la sombra, a la vislumbre, por ver si dize mejor a pelo o contrapelo o al traués, de la haz o del enués” (Magallón 1971, fol. 47 r).

Una documentación de la palabra precedida de la preposición “por” se halla en los *Coloquios de Palatino y Pinciano* de Arce de Otárola (compuestos cerca de 1550); Pinciano diserta sobre los diferentes tipos de necio que existen y afirma que “los de pie quebrado y marca menor son los que, con tres razones buenas y enteras, dicen una necedad por contrapelo, y estos son necios veniales, que con el agua bendita se libran de pena y de culpa” (Ocasar Ariza 1995, vol. 1, p. 181). Parece superfluo subrayar que todo el párrafo se basa en atribuir un sentido figurado a locuciones cuyo significado estándar era literal, como “de pie quebrado” y “de marca menor”. El significado de la expresión que nos ocupa es el que el DLE describe con el número 2: “contra el curso o modo natural de algo, violentamente”; en este caso, “contra el curso natural del argumento o de la lógica conversacional”, podríamos decir. Con esta acepción (y sin preposición alguna) encontramos documentados en el CORDE dos casos más en la primera mitad del siglo XVII y ninguno en la segunda.

Otra locución – o mejor dicho, acepción – más bien rara en el corpus es “al pelo”,³¹ la cual, respecto a la que vimos contrapuesta a “curtido” (en referencia a las pieles de algunos animales) se puede entender como su metaforización, ya que se relaciona con una sutil porción superficial de una sustancia. Como vamos a ver en el primer ejemplo, se trata de una raja finísima de pan; en el *Lazarillo de Tormes*, en efecto, el protagonista-narrador recuerda que, no pudiendo comerlos, se contentó con “dar en ellos [los panes] mil besos”; a continuación, explica: “lo más delicado que yo pude, del [pan] partido, partí un poco al pelo que él estaba, y con él pasé aquel día” (Rico 1994, p. 59). Por los mismos años, el *Manual de escribientes* de Antonio de Torquemada pone: “con mayor ánimo me atreviese a pasar a nado este golfo yendo siempre al pelo de la agua y sin samargujarme en sus honduras” (Canellada, Zamora Vicente 1970, p. 171);³² la locución se sigue

“Segunda parte de la *Crónica general de las Indias* de López de Gomara, de 1553, la *Crónica de la Nueva España* de Cervantes de Salazar, de 1560, y la mencionada traducción de Apiano. En efecto, en algunas ediciones de la obra de Apiano en español se integró la obra (o una parte) de López de Gómara y la de otros autores. Por lo tanto, no es el traductor de Apiano, sino López de Gómara el que la utiliza, aunque el CORDE no ofrece esta información.

³¹ Diferente, en cuanto al significado, de “a pelo”.

³² Aquí hay que avisar que se trata de *el pelo* y no *al pelo*, como en el ejemplo precedente; sin embargo, en ambos casos la acepción de “superficie” es la misma y en el caso del pan requiere la preposición *a*.

empleando comúnmente para indicar la superficie del agua, pero curiosamente el DLE no menciona esta acepción de la palabra, ni una eventual locución nominal.

Para terminar con las locuciones adverbiales, queda por examinar “en pelo”, que – en la acepción que vamos a comentar – aparece por primera vez en textos cuya fecha de composición es incierta, pero probablemente fueron compuestos a partir de la segunda década del siglo XVI. Citamos la oración de *Batallas y quinquagenas*, por ser una obra ya conocida: “Todo eso hoy dezir a Pero Núñez e a otros caualleros, que lo solía hazer e podía cabalgar en pelo” (Avalle-Arce 1989, p. 404). Sobra decir que se sigue empleando.

Finalmente, en una obra ya citada varias veces en estas páginas, la *Comedia llamada Floriana*, se encuentra la locución nominal “gente de pelo”, hoy desusada, que el DLE define “gente rica y acomodada” y Terreros y Pando, en su *Diccionario castellano* (1788) (lematizando “hombre de pelo”), “rico, noble”: en este sentido se emplea en dicha obra. Explica la motivación de esta locución Francisco de Osuna en su *Quinta parte del abecedario espiritual* (compuesta alrededor de 1540): “Jacob, que era pobre, no tiene pelo, y de aquí es que por injuria dezimos: Es un pelado [...]. Assí que lor ricos son pelosos y los pobres sin pelo” (Quirós García 2002, p. 527).

En la *Comedia llamada Floriana* aparecen asimismo otras expresiones como “sirvientes de pelillo”, en boca del sabio camarero de Florián: “Y por esto, con poca autoridad de los palacios, los sirvientes de pelillo, los mentirosos, chismosos, malsines, truhanes, dezidores maliciosos, chocarreros, como hallan audiencia en el señor ansí los tornan de su talle” (Canet 1996, escena XXXVII), y “servir de pellillo”, en el breve parlamento de Fulminato: “A la fe, no en balde he estado en Córdoba y hallé madre en Carmona y me llaman Fulminato. Oy en día, servir de pelillo, buena parola, facto ninguno” (Canet 1996, escena I). La regla que enuncia el personaje – y la misma elección léxica – parecen venir del italiano. La acepción del sintagma en cuestión (tanto con el sustantivo como con el verbo) es claramente negativa y se podría sintetizar como “persona o acción de baja calidad”.³³

4.2.3. *Paremiás*

Son numerosas las paremiás, sobre todo refranes, formadas alrededor de la palabra *pelo*. A lo largo de estas páginas se han citado ya varias, en particular si tenían relación con algún aspecto presente en un texto que hemos comentado, o si ellas mismas se localizaban en obras no paremiológicas; también nos hemos referido a algunas colecciones de los siglos XV y XVI. Ahora vamos a citar unas pocas más, a integración y complemento de lo que

³³ Quizás lleve algún parecido con lo que hoy significa la locución adjetiva (que sin embargo se aplica solo a los seres humanos) “de medio pelo”.

hemos venido alegando, escogiendo en particular las que contribuyen a la comprensión del contexto que la produjo.

Como se sabe, en los antiguos repertorios no se diferenciaban los refranes de otras formas fraseológicas, tanto que algunas paremias allí citadas hoy se incluirían entre las locuciones verbales, los dichos u otras formas fijas. En este apartado vamos a seguir la misma orientación, ya que no es nuestro cometido clasificar las paremias (para lo cual véanse Sevilla Muñoz, Crida Álvarez 2013). Empecemos por Vallés (1549), quien con el número 1016 cita “De que la rana tenga pelo: sereis vos bueno”, un *adynaton* que recuerda al actual “cuando la rana crie pelo”, que por supuesto significa “nunca”; Correas presenta diferentes variantes de este refrán (como, en realidad, de muchos otros); en sus orígenes, que descansan en un apólogo, se detiene también la pícara Justina en la homónima obra, como comentamos arriba. Con el núm. 2525 Vallés cita “Mudase el cielo: con el pelo”. Este refrán se encuentra en Núñez (1555) como “Mudase el celo con el pelo”, variante que da a la paremia un sentido ligado a la observación de la experiencia personal. Sin embargo, la versión de Vallés tuvo difusión, ya que alude a ella Francisco de Aldana, quien en unos versos de sus *Poesías* (1560-1578), dice “no la verán jamás mudada un pelo/ por más caras y más que mude el cielo” (Lara Garrido 1985, p. 340).

En cuanto al *Vocabulario de refranes y frases proverbiales*, Correas indaga muy a fondo el acervo popular y presenta más de 25.000 entre refranes y otros tipos de sentencias, entre los cuales: “Aunke pese a mi pelo, io seré bueno” (Combet 1967, p. 34), que alude al “mal pelo” como sinónimo de maldad. Sigue Correas: “Onbre de hecho, pelo en pecho”, dicho que él mismo explica a continuación, diciendo que “para sinifikar ke un onbre tiene valor, dezimos ke es onbre de barva, i pelo en pecho” (p. 90); este testimonio es precioso, ya que certifica la existencia de un dicho todavía vigente que no se había encontrado en el corpus. Otros refranes son chistosos, como “El rruin barvero no dexa pelo ni kuero” (p. 138). Más adelante, el lexicólogo explica que “Tener pluma se dize del ke sabe eskrivir; también se dize ‘tener pluma o pelo’, el ke es rriko” (p. 410), sintagmas que recuerdan las páginas de la *Pícara Justina* a las que nos hemos referido arriba. Otras formas fijas seguidas de aclaraciones son: “En zerro. En pelo. Dízese de una kavalgadura sin aparexos” (p. 619): también este significado es todavía vigente. “Nazer con pelo. Nazer con pluma. Dízese de los ke nazen rrikos (p. 651), que se hace eco de la locución “gente de pelo” comentada en este trabajo. Otros fraseologismos que se siguen empleando son “No tokar en el pelo de la ropa. Por: no ofender en nada” (p. 662) y “Va a pelo. Lo ke va bien. No va a pelo. Lo ke va desconforme” (p. 740), entre otros que por no detenernos demasiado no vamos a citar.

5. Conclusiones

En estas páginas hemos presentado las primeras documentaciones de uso de la palabra *pelo* en sus diferentes acepciones, y de las unidades fraseológicas cuyo núcleo era la misma palabra. *Pelo* resulta ser voz patrimonial del castellano, empleada ya desde el *Poema de mio Cid*, o sea, desde el siglo XII. A partir de ese momento, su presencia en los textos recogidos en el corpus es constante en sus múltiples significados; su aparición en diferentes tipos y géneros textuales muestra también las aplicaciones como término especializado. Lo mismo podemos decir del primer compuesto sintagmático nominal, *pelo malo*, que, combinado con un pequeño grupo de verbos sinónimos, constituye un haz locucional cuyo significado, en el transcurso de los siglos, pasa de referencial a figurado. Hay que tener en cuenta que las primeras muestras de una nueva forma fija revelan su existencia sin que podamos establecer la fecha de su verdadero nacimiento, y por otra parte los mismos textos escritos en estos siglos a menudo carecen de fecha cierta de composición. Mientras la parte atributiva de la locución cae, pues, nuevas expresiones figuradas se forman, dando vida a numerosas locuciones verbales, pero también a locuciones adverbiales, nominales y adjetivas. Por otro lado, la presencia de paremias es temprana y se debe en parte a fuentes latinas. Los refranes se insertan preferentemente en textos literarios; algunas tradiciones discursivas se adhieren a la forma dialogal que simula el habla popular y vulgar, hilvanando modismos que nos revelan las posibilidades expresivas de la lengua en determinados momentos históricos. Estos cortes sincrónicos en la diacronía son hitos en la evolución de la lengua.

Desde finales del siglo XII hasta los comienzos del XVII se han manifestado en la escritura casi todas las acepciones del vocablo *pelo* que ahora tenemos – de acuerdo con las definiciones del DLE, y con una más, no recogida en la obra académica – y muchas locuciones, cuya vida en algunos casos ha sido más breve y no ha llegado hasta nuestros días. La observación cronológica ha permitido en ocasiones llegar a vislumbrar la motivación de su creación. Un rol importante, en este sentido, es desempeñado por la lexicografía y la paremiografía, que asimismo enriquecen, aportando más variedad, el caudal fraseológico que las obras presentan.

Aunque esta indagación no sobrepasa las primeras décadas del siglo XVII, es posible afirmar que la semántica del vocablo y las pautas fraseológicas de las que disponemos en la actualidad ya quedan trazadas, gracias al temprano nacimiento del primer núcleo locucional y a la herencia latina.

Bionote: Daniela Capra es doctora en Filología Hispánica por la Universidad de Pisa y ha conseguido un *Master of Arts* en la Johns Hopkins University. Es profesora titular y ha

obtenido la acreditación a catedrática; tiene en su haber más de cien publicaciones. Ha contribuido a la fundación de dos asociaciones de lingüística y forma parte de la junta directiva de centros de investigación, revistas, colecciones, volúmenes y congresos. Posee una larga actividad de participación en conferencias y en proyectos de investigación de carácter nacional e internacional. Sus intereses científicos se dirigen hacia varios aspectos de la filología y la lingüística, en particular en perspectiva diacrónica, con estudios que se centran en facetas específicas de la textualidad o se valen de recursos digitales y *corpus* para indagar la formación de unidades fraseológicas y su desarrollo y usos en la lengua. Estudia además las relaciones culturales entre España e Italia en el Renacimiento, centrándose ya en cuestiones lingüísticas y gramaticográficas, ya culturales y traductivas.

Author's address: daniela.capra@unige.it

Bibliografía

- Alonso Cortés N. (ed.) 1969, Esteban Manuel de Villegas. *Eróticas o amatorias*, Espasa-Calpe, Madrid.
- Asociación de Academias de la Lengua Española, *Diccionario de la lengua española (DLE). Edición del Tricentenario. Actualización 2022*, en línea <https://dle.rae.es/>
- Avalle-Arce J.B. (ed.) 1989, Gonzalo Fernández de Oviedo. *Batallas y quinquagenas*, Diputación de Salamanca, Salamanca.
- Azáceta J.M. (ed.) 1956, Anónimo. *Cancionero de Juan Fernández de Íxar*, CSIC, Madrid.
- Baird Jr. H.L. (ed.) 1976, *Cuento muy feroso de Otas de Roma*, Real Academia Española, Madrid.
- Baranda C. (ed.) 1988, Feliciano de Silva. *Segunda Celestina*, Cátedra, Madrid.
- Baranda N. (ed.) 1995, Anónimo. *Historia de Enrique fijo de doña Oliva, rey de Jerusalem y emperador de Constantinopla*, Turner Libros, Madrid.
- Barrick M.E. (ed.) 1973, Gaspar Gómez de Toledo. *Tercera parte de la tragicomedia de Celestina*, University of Pennsylvania Press, Philadelphia.
- Bautista F. 2005, *La composición de la Gran Conquista de Ultramar*, en “Revista de Literatura Medieval” 17, pp. 33-70.
- Bizzarri H.Ó. 2012, *Tradición textual y tradición cultural en los Castigos del rey don Sancho IV*, en “Versants: revue Suisse des littératures romanes. Rivista svizzera delle letterature romanze. Revista suiza de literaturas románicas”, pp. 16-33. <http://www.e-periodica.ch>
- Blecua J.M. (ed.) 1971, Francisco de Quevedo y Villegas. *Jácaras*, Castalia, Madrid.
- Blecua J.M. (ed.) 1981, Juan Manuel. *Libro de la caza*, Gredos, Madrid.
- Canellada M.^a J. y Zamora Vicente A. (eds.) 1970, Antonio de Torquemada. *Manual de escribientes*, Real Academia Española, Madrid.
- Canet J.L. (ed.) 1996, Rodríguez Florián. *Comedia llamada Florinea*, en “Revista Lemir”, Valencia, edición electrónica <https://parnaseo.uv.es/lemir/textos/florinea/index.htm>
- Cantalapiedra Erostarbe D. y Moreno Uclés J. (eds.) 2004, *Introducción*, en *García de Castro. Seniloquium. Refranes que dizen los viejos*, en “Anexos de la revista Lemir”, pp. 17-26.
- Cantera Ortiz de Urbina J. y Sevilla Muñoz J. (eds.) 2002, *Los 494 refranes del Seniloquium*, Ediciones Guillermo Blázquez, Madrid.
- Cantera Ortiz de Urbina J. y Sevilla Muñoz J. (eds.) 2003, Mosén Pedro Vallés. *Libro de refranes y sentencias*, Ediciones Guillermo Blázquez, Madrid.
- Cañas J. (ed.) 1988, *Libro de Alexandre*, Cátedra, Madrid.
- Capra D. 2007, *La “Espositione” (1553) de Alfonso de Ulloa, primer glosario español-italiano*, en “Artifara” 7, Ediciones pp. 17-44.
- Capra D. (ed.) 2015, Antonfrancesco Doni. *La Zucca del Doni en español*, Accademia University Press, Torino.
- Capuano T. (ed.) 1995, Gabriel Alonso de Herrera. *Obra agricultura*, Hispanic Seminary of Medieval Studies, Madison.
- Carande R. (ed.) 1968, Anónimo. *Carta de comisión a Ribadeneira sobre los maravedies de Alonso de Palencia*, Fondo para el Fomento de la investigación en la Universidad, Sevilla.
- Carballo Picazo A. (ed.) 1953, Alonso López Pinciano. *Filosofía antigua poética*, CSIC. Instituto “Miguel de Cervantes”, Madrid.
- Cárdenas A.J. (ed.) s.f., Abraham de Toledo (Moamín). *Libro de los animales que cazan*,

- Trad. Anónima, Hispanic Seminary of Medieval Studies, Madison.
- Castañeda Delgado P. (ed.) 1994, Fray Bartolomé de las Casas. *Historia de las Indias*, Alianza Editorial, Madrid.
- Ciceri M. (ed.) 1990, Alfonso Martínez de Toledo. *Arcipreste de Talavera (Corbacho)*, Espasa-Calpe, Madrid.
- Combet L. (ed.) 1967, Gonzalo Correas. *Vocabulario de refranes y frases proverbiales*, Institut d'Études Ibériques et Ibéro-Américaines de l'Université de Bordeaux, Burdeos.
- Combet L. 1971, *Recherches sur le Refranero Castillan*, Les Belles Lettres, Paris.
- Cooper L. y Waltman F.M. (eds.) 1995, *La gran conquista de Ultramar. Ms. 187 BNM*, Hispanic Seminary of Medieval Studies, Madison.
- Corpas Pastor G. 1996, *Manual de fraseología española*, Gredos, Madrid.
- Covarrubias Orozco S. 1611, *Tesoro de la lengua castellana o española*, Luis Sánchez, Madrid.
- Crombach M. (ed.) 1971, Anónimo. *Bocados de oro*, Romanisches Seminar der Universität Bonn, Bonn.
- Crosby J.O. (ed.) 1993, Francisco de Quevedo y Villegas. *Sueño del Infierno*, Castalia, Madrid.
- Del Piero R.A. (ed.) 1971, Pedro de Veragüe. *Doctrina de la descripción o Tractado de la doctrina*, Real Academia Española, Madrid.
- Dutton B. y González Cuenca J. (eds.) 1993, AA.VV. *Poesías. Cancionero de Baena*, Visor, Madrid.
- Echenique Elizondo M.^a T. 2021, *Principios de fraseología histórica española*, Instituto Universitario “Seminario Menéndez Pidal”, Madrid.
- Echenique Elizondo M.^a T. y Pla Colomer F.P. (eds.) 2021, *DHISFRAES. Diccionario histórico fraseológico del español. Tarea lexicográfica del siglo XXI. Combinaciones de carácter locucional, prepositivo y adverbial. Muestra arquetípica*, Peter Lang, Bern/ Berlin/ Bruxelles/ New York/ Oxford/ Warszawa.
- Escudero de la Peña J.M. (ed.) 1870, Anónimo. *Relación de la dote de don Juan y la princesa Margarita*, Bibliófilos españoles, Madrid.
- Furetière A. 1690, *Dictionnaire universel, contenant generalement tous les mots françois*, chez Arnout & Reinier Leers, La Haye et Rotterdam, en línea <https://gallica.bnf.fr/ark:/12148/bpt6k3413126b>
- García Padrón D. y Batista Rodríguez J.J. 2016, *Compilación, desautomatización y desarticulación fraseológica en Quevedo*, en Echenique Elizondo M.^a T., Martínez Alcalde, M.^aJ., Sánchez Méndez J.P. y Pla Colomer F. (eds.), *Fraseología española: diacronía y codificación*, en “Anejos de la Revista de filología española” 104, CSIC, Madrid, pp. 111-131.
- García-Page Sánchez M. 1998, *Expresión fija y sinonimia*, en Wotjak G. (ed.), *Estudios de fraseología y fraseografía del español actual*, Vervuert / Iberoamericana, Frankfurt am Main / Madrid, pp. 83-95.
- Gómez Moreno Á. y Kerkhof M.P. (eds.) 1988, Marqués de Santillana (Íñigo López de Mendoza). *Sonetos al itálico modo*, Planeta, Barcelona.
- Gutiérrez-Arrese D. (ed.) 1953, Fadrique de Zúñiga y Sotomayor. *Libro de cetrería de caza del azor*, Bibliófilos españoles, Madrid.
- Herrera M.^a T. y Sánchez M.^a N. (eds.) 1999, Fray Vicente de Burgos. *Traducción de El libro de proprietatibus rerum de Bartolomé Anglicum*, Universidad de Salamanca, Salamanca.
- Herrera M.^a T. y Sánchez M.^a N. (eds.) 2000, Anónimo. *Libro de los caballos*, Universidad de Salamanca, Salamanca.

- Impr. de Aribau (ed.) 1874, Luis Milán. *El cortesano*, Madrid.
- Iradriel Murgarren P. (ed.) 1974, Anónimo. *Informe de un mercader para el fomento de las manufacturas textiles y mejora de la calidad de los paños*, Universidad de Salamanca, Salamanca.
- Kabatek J. 2017, *La construcción de la historia del español a partir de los corpus: entre “lenguas individuales” y “tradiciones discursivas”*, en Calvi M.V., Hernán-Gómez Prieto B. y Landone E. (eds.), *El español y su dinamismo: redes, irradiaciones y confluencias*, AISPI Edizioni, Roma.
- Keller, J.E. (ed.) 1961, Clemente Sánchez de Vercial. *Libro de los exemplos por A.B.C.*, CSIC, Madrid.
- Lara Garrido J. (ed.) Francisco de Aldana. *Poesías*, Cátedra, Madrid.
- Lobera F.J., Serés G., Díaz-Mas P., Mota C., Ruiz Arzálluz Í. y Rico F. (ed.) 2000, Fernando de Rojas. *La Celestina. Tragicomedia de Calisto y Melibea*, Crítica, Barcelona.
- Lozano G. (ed.) 1992, Rodrigo Fernández de Santaella. *Vocabulario eclesiástico*, Hispanic Seminary of Medieval Studies, Madison.
- Lozano López G. (ed.) 1992, Alfonso de Palencia. *Universal vocabulario en latín y en romance*, Hispanic Seminary of Medieval Studies, Madison.
- Magallón M. (ed.) 1971, Francisco Cervantes de Salazar. *Crónica de la Nueva España*, Atlas, Madrid.
- Mata Carriazo J. de (ed.) 1951, Alonso de Santa Cruz. *Crónica de los Reyes Católicos*, Escuela de Estudios Hispano Americanos de Sevilla, Sevilla.
- Meseguer Fernández J. (ed.) 1963-1964, Juan de Pineda. *Diálogos familiares de la agricultura cristiana*, Atlas, Madrid.
- Millares Carlo, A. y Artiles Rodríguez, J. (ed.) 1932, Anónimo. *Libro de Acuerdos del Concejo madrileño, 1464-1485*, Ayuntamiento de Madrid, Madrid.
- Montaner A. (ed.) 1993, Anónimo. *Poema de Mio Cid*, Crítica, Barcelona.
- Montoro del Arco E.T. 2006, *Teoría fraseológica de las “locuciones particulares”. Las locuciones prepositivas, conjuntivas y marcadoras en español*, Peter Lang, Frankfurt am Main.
- Morales García-Goyena L. (ed.) 1906-1907, Anónimo. *Real cédula de los señores Reyes Católicos*, Tipografía de López Guevara, Granada.
- Nebrija A. de 1495, *Dictionarium hispano-latinum*, s.n., Salamanca.
- Nebrija A. de 1516, *Vocabulario de romance en latín hecho por el doctísimo maestro Antonio de Nebrija nuevamente corregido y augmentado más de diez mil vocablos de los que antes solía tener*, Juan Varela de Salamanca, Sevilla.
- Núñez H. 1555, *Refranes o proverbios en romance/ que nuevamente colligiò y glossò el Comendador Hernan Nuñez*, en casa de Iuan de Canova, Salamanca.
- Ocasar Ariza J.L. (ed.) 1995, Juan de Arce de Otárola. *Coloquios de Palatino y Pinciano*, Turner, Madrid.
- Ornstein J. (ed.) 1954, Luis Lucena. *Repetición de amores*, University of North Carolina Press, Chapel Hill.
- Oudin C. 1607, *Tesoro de las dos lenguas francesa y española. Thresor des deux langues françoise et espagnolle*, Marc Orry, París.
- Palet J. 1604, *Diccionario muy copioso de lengua española y francesa [...]. Dictionaire tres ample de la langue espagnolle et françoise*, Matthieu Guillemot, París.
- Penadés Martínez I. 2006, *Sinonimia y locuciones*, en Villayandre Llamazares M. (ed.), *Actas del V Congreso de Lingüística General (León, 5-8/ 03/ 2002)*, Arco-Libros, Madrid, pp. 2217-2227.
- Pereira R. 2007, *El cuento Otas de Roma: en búsqueda de un género literario. Mémoire*

- de licence présenté à la Faculté des lettres de l'Université de Fribourg (CH)*, en línea <https://folia.unifr.ch/unifr/documents/301902>
- Pérez de Tudela Bueso J. (ed.) 1992, Gonzalo Fernández de Oviedo. *Historia general y natural de las Indias*, Atlas, Madrid.
- Pérez Priego M.Á. (ed.) 1993, Luis de Miranda. *Comedia Pródiga*, en *Cuatro comedias celestinescas*, UNED, Universidades de Valencia y Sevilla, Valencia.
- Pérez Priego M.Á. (ed.) 1994, Bartolomé de Torres Naharro. *Propalladia*, Turner, Madrid.
- Pino Jiménez J.C. (ed.) 2002, Gonzalo García de Santa María. *Traducción de la Coronica de Aragón de fray Gauberto Fabricio de Vagad*, Hispanic Seminary of Medieval Studies, Madison.
- Pujana (ed.) 1999, San Juan Bautista de la Concepción (Juan García Gómez). *Un breve tratado para los hermanos donados*, Editorial Católica, Madrid.
- Pujana J. (ed.) 2002, San Juan Bautista de la Concepción (Juan García Gómez). *Pláticas a los religiosos*, Editorial Católica, Madrid.
- Quirós García M. (ed.) 2002, Francisco de Osuna. *Quinta parte del abecedario espiritual*, FUE, Madrid.
- Real Academia Española, *Banco de datos (CORDE)*. *Corpus diacrónico del español*, en línea <https://corpus.rae.es/citar.htm>
- Real Academia Española, *Nuevo tesoro lexicográfico de la lengua española (NTLLE)*, en línea <https://www.rae.es/obras-academicas/diccionarios/nuevo-tesoro-lexicografico-0>
- Regalado López T. (ed.) 2000, Juan de Alcega. *Libro de geometría, práctica y traza*, CILUS, Salamanca.
- Rey Hazas A. (ed.) 1977, Francisco López de Úbeda. *La pícaro Justina*, Editorial Nacional, Madrid.
- Rico F. (ed.) 1994, Anónimo. *Lazarillo de Tormes*, Cátedra, Madrid.
- Rodríguez Herrero Á. (ed.) 1959, Andrés de Poza. *De la antigua lengua, poblaciones y comarcas de las Españas*, Ediciones Minotauro, Madrid.
- Rodríguez Molina J. y Octavio de Toledo y Huerta Á. 2017, *La imprescindible distinción entre texto y testimonio: el CORDE y los criterios de fiabilidad lingüística*, en “Scriptum digital” 6, pp. 5-68.
- Rodríguez Puértolas J. (ed.) 1968, Fray Íñigo de Mendoza. *Cancionero*, Espasa-Calpe, Madrid.
- Rojas Garcidueñas J. (ed.) 1958, Fernán González de Eslava. *Coloquio tercero a la consagración del doctor don Pedro Moya*, Porrúa, México.
- Rubio Pardo C., Moreno Valcárcel T. y Fuente C. de la (eds.) 1979, Anónimo. *Libro de acuerdos del Concejo madrileño*, Ayuntamiento de Madrid, Madrid.
- Sagüés Azcona P. (ed.) 1951, Fray Diego de Estella. *Modo de predicar y modus concionandi*, CSIC, Madrid.
- San Vicente Santiago F. 2022, *El diccionario bilingüe de italiano y español fuente documental de la lengua española*, en *Saberes humanísticos, ciencia y tecnología en la investigación y la didáctica del hispanismo*, pp. 310-328.
- Sánchez-Prieto P. (ed.) 1999, Anónimo. *Testamento [Documentos del Archivo Histórico Nacional (a 1200-a 1492)]*, Universidad de Alcalá, Madrid.
- Sánchez-Prieto Borja P. (ed.) 2004, *Fuero General de Navarra [Versión A] BNM Ms. 17653*, Universidad de Alcalá de Henares, Alcalá de Henares.
- Sendra Cendra M^aD. (ed.) 1966, *Relación de los productos del peaje de Alagón [Aranceles aduaneros de la corona de Aragón]*, Anubar, Valencia.
- Sevilla F. (ed.) 2001, Mateo Luján de Saavedra (Juan Martí). *Segunda parte de la vida del pícaro Guzmán de Alfarache*, Castalia, Madrid.

- Sevilla Muñoz J. y Crida Álvarez C.A. 2013, *Las paremias y su clasificación*, en “Paremia” 22, pp. 105-114.
- Temprano J.C. (ed.) 1995, Anónimo. *Proverbios. Moñino 73*, Hispanic Seminary of Medieval Studies, Madison.
- Vallés P. 1549, *Libro de refranes copilados por el orden del A.B.C.*, en casa de Juana Millán, Çaragoça.
- Varela C. y Gil J. (eds.) 1992, Anónimo. *Diario del primer viaje de Colón [Textos y documentos completos de Cristóbal Colón] 1492-1493*, Alianza, Madrid.
- Vicente Llavata S. 2011, *Estudio de las locuciones en la obra literaria de don Íñigo López de Mendoza (Marqués de Santillana). Hacia una fraseología histórica del español*, Anejo n. 76 de la revista “Quaderns de Filologia”, Universitat de València, Valencia.

EXPLORER LES MOTS DU POLITIQUE DANS LA TRANSFORMATION NUMERIQUE

Analyser le lexique politique dans des contextes et selon des ressources en évolution

JULIEN LONGHI

CY CERGY PARIS UNIVERSITÉ, AGORA LAB, INSTITUTE OF DIGITAL HUMANITIES

Abstract – The evolution of digital technologies has profoundly transformed the way we exchange political ideas. The rise of social networks, blogs, and online discussion forums has provided an increasingly accessible platform for political organizations to communicate and interact with voters. In order to explore words in digital transformation, and the use of tools and/or approaches for studying lexicon and phraseology in evolving discursive domains, we chose to focus our study on words in the political domain, in the context of the 2017 and 2022 French political elections. Thus, in the interval of two presidential campaigns, the political context has changed enormously (evolution of the French political landscape, recomposition of parties and the electorate), and corpus analysis technologies have also undergone a great evolution. The analysis of digital political discourse has become increasingly important due to the growing importance of the Internet and social media in public debate and opinion formation. New challenges in digital political discourse analysis include the sheer volume of data (which is often noisy and contains redundant or irrelevant information), polarization, misinformation, and the difficulty of distinguishing between facts, opinions, and rumors, especially in short messages. This paper will therefore address both the methodological and technological transformations, as well as the discursive and argumentative transformations, of the analysis of political words in an electoral context, through the presentation of different projects and initiatives that have marked the scientific landscape during the campaigns. Finally, to deepen this inventory, and to address the issue of lexicon and political discourse, we will focus on the theme of the analysis of the candidates' style, by highlighting the way in which the use of deep learning and textual statistics can help to better understand the evolution of political discourse, and to measure the contribution of recent technologies and tools mobilizing Artificial Intelligence.

Keywords: political lexicon; textual statistics; digital corpus; artificial intelligence; style.

Le recours à l'ordinateur, rendu nécessaire à la fois par l'ampleur des opérations de dépouillement envisagées et par le volume des calculs statistiques couramment effectués, implique en retour que l'on définisse de manière toujours plus précise l'ensemble des règles qui présideront au dépouillement automatique des textes stockés en machine.

(L. Lebart & A. Salem « Statistique Textuelle », 1994, p. 18).

1. L'analyste du discours face aux évolutions linguistiques et technologiques : enjeux et perspectives

Afin de rendre compte, de manière linguistique, des domaines discursifs en évolution, nous avons choisi de concentrer notre étude sur les mots du domaine politique, dans le cadre des élections politiques françaises de 2017 et 2022. Ainsi, dans l'intervalle de deux campagnes présidentielles, le contexte politique a énormément changé (évolution du paysage politique français, recomposition des partis et de l'électorat), et les technologies d'analyse de corpus ont également connu une évolution spectaculaire. La littérature scientifique en analyse du discours (notamment) témoigne de ces évolutions, tout comme différents projets de recherche qui ont, lors de ces élections, proposé des analyses et produit des articles ou des ouvrages.

Cet article abordera donc à la fois les transformations méthodologiques, technologiques, mais aussi discursives et argumentatives, de l'analyse des mots politiques en contexte électoral (à la suite de Longhi 2016 et 2020), à travers la présentation de deux projets et initiatives qui ont marqué l'actualité scientifique lors des campagnes. Cela permettra de présenter à la fois les enjeux sociétaux et informationnels de tels projets, mais surtout de décrire précisément les outils et technologies mobilisées, et leur évolution sur la période 2017-2022. Nous présenterons ensuite deux projets menés à CY Cergy Paris université, qui s'intègrent dans le mouvement des deux précédemment décrits : #Idéo2017 et PolitiQuiz. Enfin, pour approfondir cet inventaire, et aborder la problématique du lexique et du discours politique, nous approfondirons le thème de l'analyse du style des candidat-e-s, en mettant en valeur la manière dont l'usage du *deep learning* et des statistiques textuelles peuvent aider à mieux comprendre l'évolution du discours politique, et permettre de mesurer l'apport des récentes technologies et des outils mobilisant l'Intelligence Artificielle.

2. Des projets emblématiques d'analyse du discours politique : objectifs et assise scientifique

2.1. *Politoscope et Gargantext : analyse communicationnelle et politique*

Le Projet Politoscope, CNRS Institut des Systèmes Complexes Paris Ile-de-France (ISC-PIF), <http://politoscope.org> aborde des questions politiques d'un point de vue informatique et communicationnel, mais pose des questions tout à fait utiles pour des travaux en analyse du discours : ces chercheurs questionnent en effet la manière dont les différentes communautés politiques s'organisent sur Twitter (devenu depuis X¹) pour diffuser de l'information, ils recensent les thèmes de prédilection des candidats, et leur appropriation par les communautés, ils observent la création et la propagation des infos/fake news, et cherchent à cartographier l'importance des candidats.

Pour l'analyse textuelle, le Politoscope utilise Gargantext, défini comme un « outil de text-mining » : en contexte électoral, les équipes ont « analysé les mesures proposées par l'ensemble des candidats ainsi que tous leurs tweets captés par le Politoscope » afin de « délimiter un vocabulaire spécifique au discours politique [...] dont les termes ont été catégorisés en grands thèmes de manière semi-automatique », même si les auteurs reconnaissent quelques erreurs de classification. L'exemple donné sur le site de Gargantext permettra de clarifier l'approche :

Pour ne prendre qu'un exemple, sous la catégorie Économie & fiscalité sont regroupés tous les tweets qui mentionnent au moins une fois l'un des termes suivants : entreprise*; pme; pmi; actionnaire*; relocalisation*; crédit impôt; nationalisation*; commerçant*; plus-values; délocalisation*; cotisations sociales; scop; niches fiscales; gafa; cice; taux d'imposition; retenue à la source; crédit impôt-recherche; impôt sur les sociétés; austérité; fisc*; impôt; imposition; cotisations sociales; retenue à la source; etc. Lors de la présidentielle 2017, dans un souci d'équité et afin d'être le plus exhaustif possible, nous avons invité le 28 février 2017 toutes les équipes de campagne à nous communiquer des éléments concernant leur programme et les mot-clés qu'ils souhaitaient mettre en avant.

On comprend donc que des catégories/thèmes de campagne ont été définis, et qu'ils sont caractérisés par un vocabulaire, qui permet ensuite de décider la thématique d'un tweet. Gaumont *et al.* (2018, pp. 10-11) précisent la méthode :

¹ Nous conservons dans cet article le nom Twitter quand il s'agit de la plateforme telle qu'elle était considérée/analysée avant le changement de nom, le 23 juillet 2023. Le terme tweet(s) est conservé, indépendamment de la période, pour désigner les messages produits sur Twitter/X.

Pour mieux caractériser les communautés politiques, reconstruites uniquement à partir de l'analyse des interactions sociales, nous avons analysé leur profil sémantique. Notre objectif n'est pas de les caractériser de manière exhaustive, mais de donner un aperçu de la façon dont les communautés diffèrent dans leur agenda politique et dans leur manière de contextualiser l'information. Pour ce faire, nous avons effectué une analyse du discours politique afin d'extraire les principaux débats de cette campagne. Nous avons utilisé pour cela deux types de sources : les mesures politiques rapportées dans les programmes de campagne, écrites dans un langage contrôlé, et les tweets des candidats, qui empruntent souvent au langage parlé. [...] nous avons analysé ces deux corpus à l'aide du logiciel de fouille de texte du CNRS Gargantext (<https://gargantext.org> ; open source à <https://github.com/ISCP/IF/gargantext>), qui met en œuvre des méthodes avancées de traitement automatique du langage et de détection de thématiques par l'analyse des co-termes ; ainsi que des méthodes de visualisation pour la représentation des résultats et la navigation interactive. Dans les premières étapes, Gargantext utilise une combinaison de lemmatisation, de post-tagging, d'analyse statistique comme tf-idf [47] et d'analyse des relations de généralité entre termes [48] pour identifier dans le texte quelques milliers de mots clés qui sont spécifiques au discours politique. Ces mots-clés ont ensuite été triés par les 377 auteurs afin de sélectionner les plus significatifs (i.e. les mots vides ou les expressions mal formées qui n'auraient pas été filtrées lors des étapes de text-mining ont été supprimés, des hashtags ou néologismes importants de Twitter comme *frexit* ont été ajoutés). Enfin, toutes les mesures politiques ont été relues attentivement avec les mots-clés sélectionnés mis en évidence dans le texte afin de vérifier qu'il ne manquait aucun mot-clé important. Cela a conduit à un vocabulaire de près de 1600 groupes de mots-clés qualifiant les thèmes de la campagne présidentielle (voir Texte I dans S1 File pour la liste des mots clés). Nous avons utilisé la mesure de proximité confiance pour évaluer la proximité thématique entre les termes sélectionnés. Cette mesure est le maximum entre deux probabilités conditionnelles. [...] Nous avons appliqué l'algorithme de Louvain [35] pour identifier des groupes de termes délimitant des sujets. Enfin, nous avons généré la carte thématique pour chacun de ces deux corpus.

Dans cette méthode (tout comme la LDA, non utilisée ici car peu adaptée aux formats courts comme les tweets) « les sujets sont définis comme des “sacs de mots”, déduits des statistiques d'apparition au sein des documents de termes issus d'une liste de mots clés définie à l'avance », et même si les mots sont ensuite liés aux thèmes, l'analyse reste tributaire d'une vision atomisée du lexique, qui ne prend pas en compte la phraséologie, la syntaxe, voire la dimension argumentative des termes. En outre, cette approche se distingue des méthodes textométriques dans lesquelles « le texte est caractérisé par ses mots par rapport à leur usage dans le corpus, le mot est caractérisé par ses cooccurrents, etc. » (Pincemin 2012). Cela n'est pas l'objectif de ce projet, qui est centrée sur l'analyse des communautés politiques, et la circulation des messages, et il est néanmoins très intéressant pour décrire l'influence des discours numériques sur les dynamiques politiques. Ce projet est donc

remarquable par son analyse topologique de la twittosphère, et sa mesure des traces numériques et des dynamiques des communautés. Il connaît néanmoins certaines limites si on le considère avec un prisme discursif, puisque le contenu des messages est analysé de manière assez systématique et peu contextuelle.

Un autre projet, porté par une équipe de linguistes-informaticiens, et d'historiens, s'inscrit dans l'objectif de décrire la campagne lors de son déroulement, avec une attention plus spécifique à la matérialité des corpus.

2.2. Mesure du discours et Hyperdeep : analyse de corpus et du vocabulaire

Comme indiqué sur le compte X du projet, <http://mesure-du-discours.unice.fr> est une plateforme web d'analyses statistiques de corpus politiques issue des travaux de recherches de l'UMR 7320 – BCL. Le projet combine des méthodes logométriques et le *deep learning*. Il croise donc fortement les intérêts et objectifs des projets menés à CY Cergy Paris université que nous décrirons dans un second temps.

Le site ne documente pas beaucoup son objectif général, mais le site de la bibliothèque universitaire de l'UCO (<https://bu.uco.fr/ressource/mesure-du-discours>) explicite les enjeux :

Ce site web est co-développé par l'équipe "Logométrie" du laboratoire "Bases, Corpus, Langage" de l'Université de Nice et l'équipe SPARKS du laboratoire I3S localisé dans la même ville. Plus qu'un site, c'est un outil d'analyse des discours politiques français fonctionnant grâce à l'intelligence artificielle et la statistique textuelle.

[...]

On y trouve les discours produits par les présidents successifs de la 5e République, ceux produits lors des dernières élections présidentielles (2007, 2012 et 2017) et plusieurs discours révolutionnaires (1789-1799). Chaque discours est accessible en texte intégral auquel s'ajoutent différents résultats d'analyse textuelle : les mots les plus spécifiques employés, les passages clés, les distributions statistiques de mots, les cooccurrences ou encore l'analyse par "deep learning" de l'orientation politique des discours.

L'essentiel de la base méthodologique et théorique de ce site est présenté dans l'ouvrage *L'Intelligence artificielle des textes. Des algorithmes à l'interprétation* (2021) publié sous la direction de Damon Mayaffre et Laurent Vanni. Plusieurs chapitres détaillent les aspects principaux et novateurs de l'approche. Dans leur article intitulé « Deep learning et description des textes. Architecture méthodologique », Laurent Vanni et Frédéric Precioso expliquent que « le deep learning montre aujourd'hui de nouveaux atouts et bénéfiques pour l'Analyse de Données Textuelles (ADT) »

(p. 15), notamment parce que la boîte noire que constitue souvent l'IA « s'ouvre peu à peu et libère une connaissance inédite des textes », complétant ainsi les méthodes traditionnelles et « offrant un champ de vision plus large et un parcours interprétatif plus fin ». La réflexion autour de l'IA concerne à la fois la nature des observables, et le statut accordé au texte et plus largement au langage (p. 17) :

Avec les réseaux de neurones, il est possible d'observer des zones descriptives du texte plus ou moins longues, possiblement discontinues, définies par l'apprentissage du modèle. En deep learning il y a peu d'a priori sur le texte et ses observables : les mots, les lemmes, les codes grammaticaux sont encodés en valeurs numériques et le modèle recherche dans cette masse d'informations tout élément susceptible de lui permettre de converger vers une solution optimale.

L'encodage des mots dans les réseaux de neurones passe par le recours aux *embeddings* qui « à chaque mot un vecteur (potentiellement de grande dimension) de telle manière que la distance euclidienne des vecteurs reflète, sinon directement le sens des mots, toutefois leurs rapports et organisation sémantiques » (p. 22) : ceci offre en quelques sortes un changement de paradigme, puisque

l'approche fréquentielle statistique classique donne aux mots une valeur commune pour toutes ses occurrences, quels que soient ses contextes. [...] Avec la convolution, l'approche passe du fréquentiel au séquentiel, chaque mot est pris dans son contexte et engendre une représentation vectorielle particulière liée au contexte. Même si le fréquentiel joue un rôle important dans l'apprentissage, cette fenêtre contextuelle coulissante garantit une analyse séquentielle et contextualisante des mots. (Vanni, Precioso 2021, p. 29)

Plus généralement, pour la problématique de l'article et de ce numéro, on notera un changement radical de la manière de concevoir les mots, qui sont « convertis » en vecteurs, et deviennent donc des données numériques, qui peuvent ensuite être traitées par des algorithmes.

Plus précisément, les auteurs expliquent (p. 33) que « pour obtenir une représentation pertinente des mots, une des solutions les mieux reconnues par la communauté est d'utiliser une méthode appelée Word2Vec, sorte de préapprentissage qui utilise différents modèles possibles comme CBOW ou Skip-Gram de (Mikolov *et al.* 2013). Cette méthode efficace prend en compte l'axe syntagmatique et l'axe paradigmatisque [et] donne un moyen efficace de fixer la représentation des mots avec un état compréhensible pour l'homme basé sur la cooccurrence des mots », avec finalement (p. 34) une « représentation des mots par leurs profils cooccurentiels ». Pour pouvoir aboutir sur une forme interprétable pour l'humain, les auteurs introduisent une notion supplémentaire à leur modèle, la *déconvolution* : cette méthode

« agit comme un décodeur pour les réseaux à convolution. Adaptée au texte cette déconvolution propose une nouvelle manière de lire les textes sous une forme abstraite proposée par le deep learning » (p. 41). Les auteurs appellent ce processus « text deconvolution saliency (TDS) » qu'ils définissent comme « un ensemble de processus algorithmiques qui visent à inverser les effets de la convolution » (p. 41) : ils expliquent également que « linguistiquement cela se traduit par une mise en valeur de certains mots, expressions ou patterns que le réseau utilise pour prendre sa décision (classification) »² (p. 46).

Le chapitre suivant, intitulé « Littérature et intelligence artificielle » et écrit par Étienne Brunet, Ludovic Lebart, Laurent Vanni, évoque des enjeux plus généraux du recours en IA dans l'analyse des textes, et détaillent surtout le type de rapport qui est donné à l'analyste avec les résultats (pp. 74-75) :

lorsque les techniques de l'Intelligence artificielle sont mises en œuvre, la reconnaissance (ou identification) n'est pas absolue. Le processus de classification aboutit à un continuum de valeurs échelonnées de 0 à 100. Cette approximation en pourcentages est assez habituelle dans les conclusions statistiques, et particulièrement dans les méthodes éprouvées de la lexicométrie. Mais il faut ici souligner une différence essentielle dans l'approche des faits. L'habitude s'est imposée dans les études lexicométriques de constituer un corpus en rassemblant des textes que l'on oppose les uns aux autres en s'aidant de la « norme » interne représentée par l'ensemble du corpus. Il y a certes des avantages pratiques à procéder de la sorte mais cela ne va pas sans bousculer la vraie démarche statistique qui exclut radicalement l'échantillon de la population. Pour savoir si un échantillon (ou un texte) peut ou non appartenir à la population (ou corpus), les tests autorisés n'admettent pas sa présence, même proportionnellement faible, dans les données de référence, dites aussi données d'apprentissage. La procédure que suit en particulier le deep learning respecte intégralement cette distinction, et dénonce immédiatement les forfaitures, en parlant de façon assez dérogatoire d'apprentissage par cœur (overfitting). Quand le texte à examiner figure par mégarde ou ignorance dans les données d'apprentissage, un seuil scandaleux de reconnaissance signale le non-respect du principe de séparation. Dans la même situation, la lexicométrie traditionnelle enregistre peu de différence dans les résultats, que le texte à examiner soit extérieur ou non au corpus. On peut en conclure a priori que l'approche du deep learning, étant plus exigeante et plus sévère dans les conditions de traitement, laisse espérer une sensibilité plus grande et plus exacte à la spécificité des textes.

² Ou encore : « Le TDS est donc un nouvel outil qui attribue un poids à chaque mot, basé sur un calcul non linéaire (lié aux fonctions d'activation de chaque neurone, elles-mêmes non linéaires) et que la statistique classique ne peut engendrer. Avec la convolution, une même forme graphique prend un poids différent suivant son contexte, là où la statistique ne propose qu'un seul score de spécificité (ou z-score) pour chacune des classes ou qu'un profil cooccurentiel général unique dans le corpus ».

On reconnaît donc une complémentarité des approches, mais également une différence de fond en ce qui concerne la reconnaissance de manière « non absolue » en pourcentage (pour le *deep learning*) d'un côté, ou la proximité/distance des textes en fonction de la norme établie par l'ensemble des textes (le corpus, dans la lexicométrie/statistique textuelle).

Ces technologies sont ensuite mobilisées en ligne, comme dans la Figure 1 avec l'analyse du mot « combat » dans l'élection présidentielle 2017 :

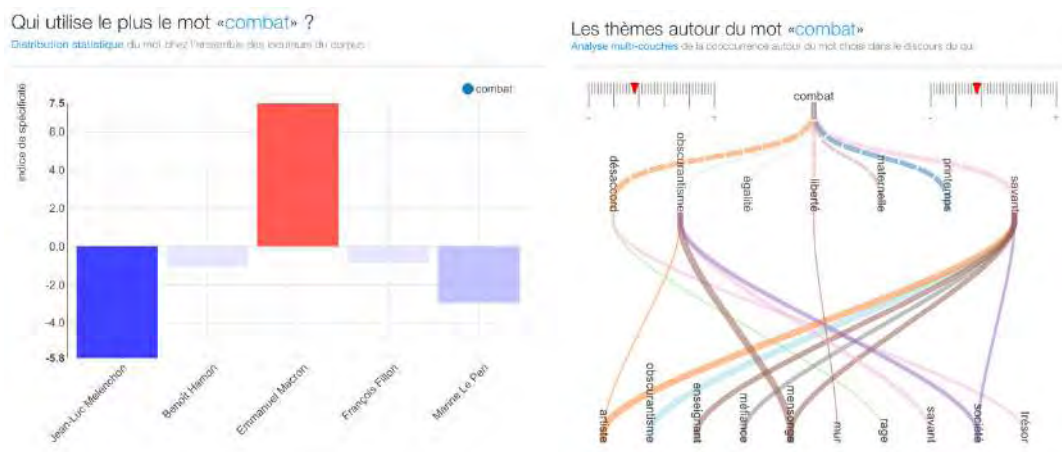


Figure 1

Analyse du mot « combat » dans Mesure du Discours (présidentielle 2017).

Les préoccupations de ce site rejoignent un certain nombre des enjeux de projets que nous avons également menés, centrés en particulier sur les liens entre lexique, idéologies, et argumentation politique.

3. #Idéo2017 et PolitiQuiz, deux projets ouverts et interactifs pour explorer les mots du politique dans la transformation numérique

À partir du second semestre 2016, et dans le cadre d'un projet financé par CY Fondation, nous avons mis en place un projet de recherche dont l'objectif était de rendre les outils et techniques de l'analyse du discours numérique au plus grand nombre.

3.1. Interactivité et acquisition de nouvelles connaissances

Le projet #Idéo2017, centré sur les élections présidentielles 2017 a permis la mise en ligne d'une plate-forme d'analyse, en temps réel, des tweets des candidats, avec plusieurs types de métriques et visualisations de leurs corpus. La chaîne de traitement proposée était celle présentée en Figure 2 :

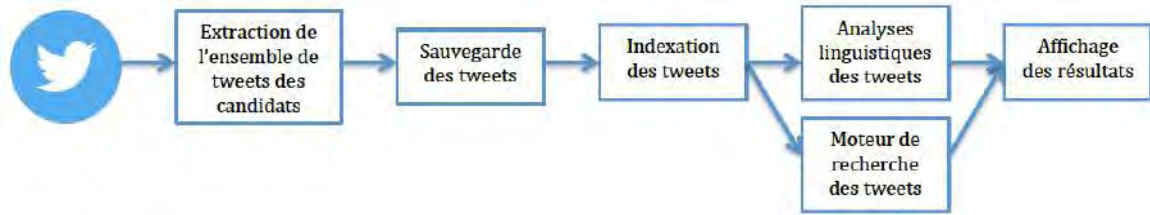


Figure 2
Chaîne de traitement de la plateforme #Idéo2017.

En ligne, l'interface visible par l'utilisateur se composait de trois parties, permettant de lancer des analyses linguistiques de mots, de corpus des candidats, ou d'accéder à un moteur de recherche (Figure 3) :



Figure 3
Plateforme #Idéo2017 côté utilisateur.

Plusieurs publications ont déjà résumé les enjeux et concepts du projet #Idéo2017, notamment dans Longhi (2019) :

L'analyse "J'analyse les tweets qui contiennent le mot..." permet à l'utilisateur de choisir un mot parmi les 13 mots qui sont souvent employés dans les débats politiques (ALDUY 2017). Cette entrée donne accès à quatre analyses possibles : l'usage de ce mot par les différents candidats (sur/sous-emploi et fréquences de la forme exacte), les mots associés à ce mot (analyse de similitudes, basée sur les cooccurrences entre les mots), l'emploi de ce mot et ses dérivés par les différents candidats [...] et le nuage de mots. Ces analyses sont en fait des résultats produits grâce au code du logiciel d'analyse textuelle Iramuteq, issus de calculs qui portent dans le logiciel des noms plus techniques (l'analyse de similitude devient par exemple les mots associés à ce mot).
L'analyse "J'analyse les tweets de [candidat]" permet à l'utilisateur de choisir un candidat parmi les 11 candidats (ou le corpus global des 11 candidats) afin d'analyser ses tweets via les techniques suivantes : les mots les plus utilisés, les thématiques (issues de la méthode Reneirt), les relations entre les mots, le

nuage de mots, les spécificités des différents candidats (possible si l'utilisateur a choisi d'analyser tous les candidats en même temps).

Enfin, le moteur de recherche permet à l'utilisateur de faire des recherches sur toute la base des tweets, grâce à un outil appelé ElasticSearch.

Ces analyses, comme dans *Mesure du discours*, permettent à l'utilisateur d'accéder à des résultats d'analyses textométriques, et de les rendre visibles dans le cadre d'une interaction avec l'interface. Il est ici davantage encore question de centrer la plateforme sur des « objets discursifs » (Longhi 2008, 2015) dans le but de pouvoir comparer la dimension idéologique des discours des candidats, et le travail réalisé sur la matérialité des discours qu'ils tiennent.

Dans ce cadre du projet *PolitiQuiz*³, nous avons souhaité contribuer à la compréhension des élections régionales 2021 en Île-de-France, puis aux élections présidentielles 2022 en France. Ce projet propose des questions et analyses permettant de mieux connaître les discours des candidates, leur style, leurs thèmes, et leurs programmes. Mais au-delà du simple quiz, *PolitiQuiz* offre aux utilisateurs l'accès aux technologies développées dans le cadre de l'institut des humanités numériques (IDHN) de CY.

3.2. Des questions pour mieux comprendre les discours des candidats

L'application *PolitiQuiz* proposait initialement trois grandes thématiques : les connaissances politiques, l'attribution d'auteur, et les grandes thématiques de campagnes (sécurité, santé, etc.). Dans la version de 2022, le choix a été simplifié à deux catégories, davantage centrées sur la thématique politique que sur le type de traitement. On distinguait ainsi la « culture générale » d'une partie concernant « Les programmes en question » (voire Figure 4) :

³ Plusieurs participant-e-s ont collaboré au projet *PolitiQuiz*, dans ses différentes phases : Jérémie Demange, Ingénieur d'études (IGE), a été le maillon central du projet sur le plan technique et multimédia. Des stagiaires (de DUT ou de Master) ont été impliqués dans la première phase (design, questions) ; des doctorantes (Lise Pernet et Rose Moreau Ragueneau) ont également participé à la 2nde phase (élection 2022) en travaillant notamment sur les programmes et sur la dimension participative. Cette équipe a donc participé à ce qui est résumé dans les sections 3.2 et 3.3.

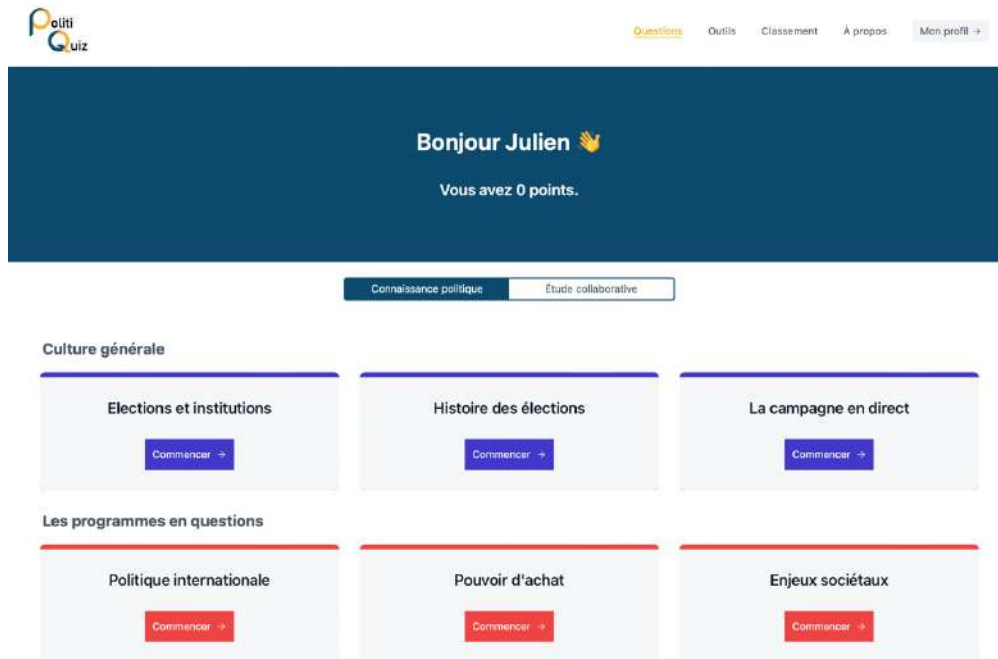


Figure 4
Plateforme PolitiQuiz côté utilisateur.

Les questions ont été produites à partir de l'analyse des tweets des principaux candidats à l'élection (régionale puis présidentielle). Pour cela, nous avons utilisé différents outils ou logiciels (comme Iramuteq ou Tropes) et méthodes d'analyses afin de pouvoir créer de nouvelles questions. Il s'agissait en effet d'extraire des caractéristiques thématiques, sémantiques, du corpus de référence initialement constitué, pour produire des questions adaptées à l'élection en cours. Par exemple, en utilisant le « calcul de spécificité » du logiciel Iramuteq, nous avons mis en valeur les mots sur- et sous-employés par tel ou tel candidat. À partir de ces résultats, des questions ont été rédigées (par l'équipe du projet et une stagiaire recrutée pour chacune des élections), pour permettre aux utilisateurs de tester leurs connaissances sur les mots les plus utilisés par les différents candidats, ou de comparer les discours des candidats, et donc leurs propositions, comme la Figure 5 s'illustre :



Figure 5
Exemple de question posée sur PolitiQuiz.

Le travail a été réalisé de manière collaborative comme indiqué en Figure 6 :

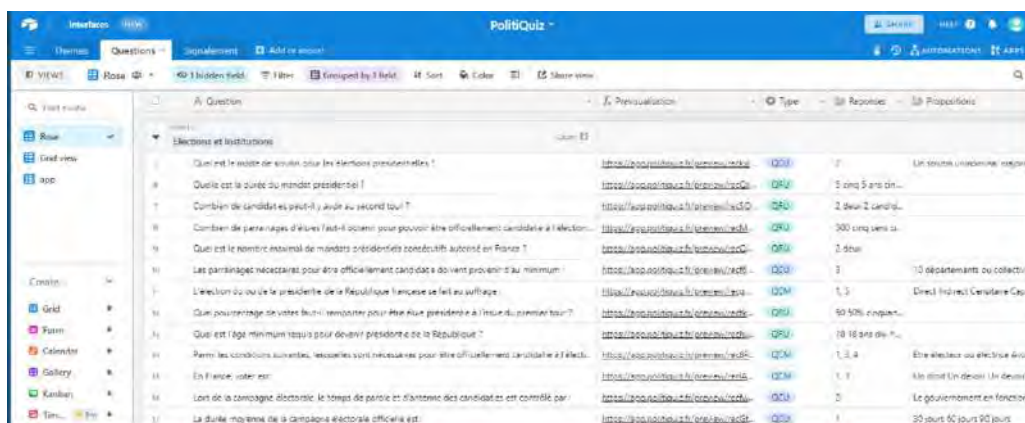


Figure 6
Document collaboratif de saisie des questions
(complété par les stagiaires et participant-e-s au projet).

Notons que l'application possédait un algorithme auto-adaptatif, qui adaptait les questions en fonction des réponses et proposait ensuite des questions plus ou moins complexes. En fonction des questions, nous avons attribué à chacune d'elle un niveau de difficulté sur une échelle de un à trois. En fonction de la progression de l'utilisateur, l'outil proposait une pause d'un ou deux jours pour faire travailler la mémoire, puis de nouvelles questions inédites apparaissaient à nouveau sur l'application.

Cette dimension pédagogique vient du fait que le projet PolitiQuiz avait des liens avec un autre projet Open Source d'une start-up d'État : Pix.fr. Le code de Pix a été redéveloppé (par JérémY Demange) en ajoutant des questions, un classement et une interface aux couleurs du PolitiQuiz. L'application avec ces modifications permettait d'offrir une interface intuitive et facile à prendre en main, qui embarque un certain nombre de technologies (Figure 7) :



Figure 7
Page de création de compte dans PolitQuiz.

Le site s'appuyait sur différentes innovations scientifiques, en les rendant accessibles, et ludiques, par exemple :

- pour les questions de la catégorie « attribution d'auteur », l'algorithme d'attribution d'auteur développé dans le cadre du projet CHEMI « IRITA », a été combiné à un processus de génération de texte par l'usage du modèle GPT-Neo ;
- pour les questions liées aux thématiques politiques, nous avons préalablement eu recours aux classifications statistiques des corpus. En effet, un traitement d'un large corpus de tweets des candidats a été établi avec le logiciel de statistique textuelle Iramuteq, afin de caractériser les mondes lexicaux (classes) qui rendent compte des grandes thématiques : éducation, sécurité, etc.

Une page de classement est également accessible, permettant de comparer ses résultats avec ceux des autres utilisateurs de l'application. En utilisant le principe du jeu, du classement, voire du concours, nous espérons sensibiliser les électeurs aux différents enjeux du scrutin, tout en leur offrant des connaissances et informations, de manière interactive et ludique.

3.3. Science participative et linguistique populaire : l'analyse des nominations

En plus de la dimension ludique et analytique, nous avons, à la fin de l'élection 2022, pris l'initiative d'introduire une nouvelle composante, intitulée « Étude collaborative » (figure 8) :

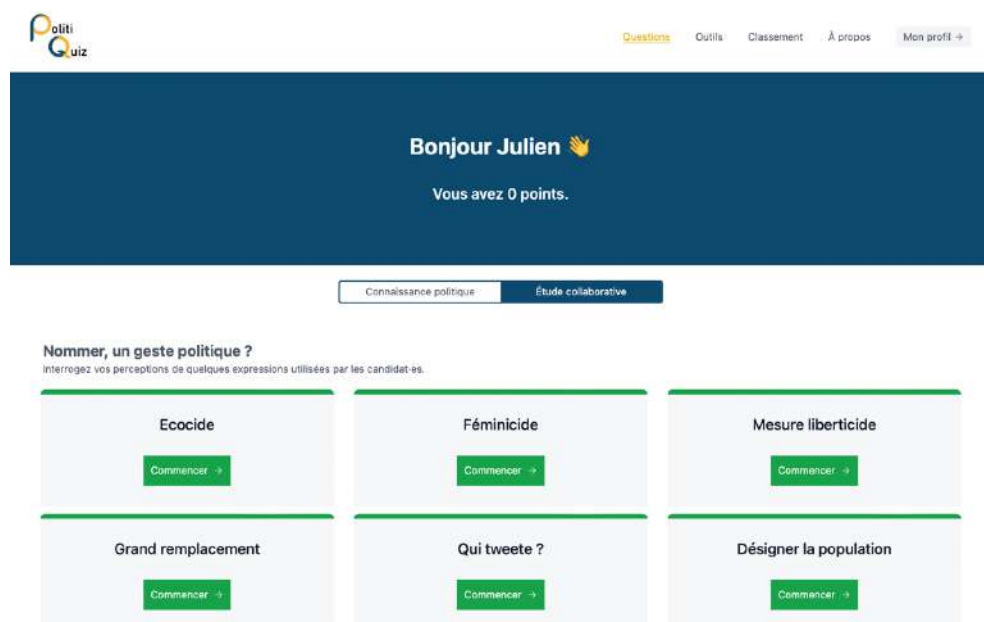


Figure 8

Étude collaborative au sujet de la nomination (« Nommer, un geste politique » ?).

Il s'agissait dans cette rubrique d'« interroger [les] perceptions de quelques expressions utilisées par les candidat.e.s ». L'intitulé, « Nommer, un geste politique », inscrit ce questionnement dans le domaine de la nomination et de l'acte de nommer, qui constitue une voie d'entrée privilégiée dans l'analyse sémantiques des discours. Des termes comme « écocide », « féminicide », ou des expressions comme « Grand remplacement », étaient questionnées.

4. Articuler « recherche-action » et réflexivité : réinterroge le concept de style au prisme des formations discursives et des genres de discours

La réalisation de tels projet, en prise direct sur la campagne, nécessite la production de résultats rapides, et la mise en évidence de tendances sans forcément pouvoir les approfondir sur le moment. Ces corpus, résultats préliminaires, et analyses, sont autant d'hypothèses qui peuvent ensuite être approfondies par le chercheur. En particulier, notre entrée lexicale dans l'analyse du discours politique permet de comparer le vocabulaire des candidats, mais aussi leur style.

4.1. Comparaison des spécificités : le vocabulaire comme entrée dans le style en politique

Lors de ces deux élections, certain-e-s même candidat-e-s ont été présents, et cela nous a invité à questionner la permanence du style de ces personnalités, ainsi que la cohérence de leur discours à cinq années d'intervalle. La question du contexte est également à considérer, puisque les enjeux peuvent évoluer d'une élection à l'autre. Nous avons donc choisi de comparer trois personnalités, lors de ces deux élections : Emmanuel Macron, Marine Le Pen et Jean-Luc Mélenchon. Emmanuel Macron étant président en 2022, sa communication diffère de celle d'une communication de campagne classique, ce qui nous a conduit à adapter la collecte comme indiqué sur la figure 9 :

	auteur	size
0	EM1	2612
1	EM2	298
2	JLM1	3722
3	JLM2	1000
4	MLP1	2440
5	MLP2	1000

Figure 9
Composition du corpus.

Nous n'avons que 298 tweets d'Emmanuel Macron sur cette période, et nous avons ensuite limité à 1000 le nombre de tweets de Jean-Luc Mélenchon et Marine Le Pen. Le recours au logiciel Iramuteq est utile ici, puisqu'il permet d'encoder des variables, comme l'auteur ou la date, et de calculer ce qui est spécifique de telle ou telle partie de corpus (figure 10) :



Figure 10
Spécificités analysables dans Iramuteq.

Ceci est rendu possible par l'encodage du corpus (Figure 11), où on peut d'ailleurs distinguer l'auteur (Emmanuel Macron par exemple) du candidat spécifique (appelé ci-dessous spécificité, « EM2 » soit Emmanuel Macron de 2022) :

```
**** *auteur_EM *date_24_03_2022 *spécificité_EM2
Pour éviter une crise alimentaire, il nous faut réagir. J'ai souhaité lancer, en lien
direct avec l'Union africaine, l'initiative FARM : Food and Agriculture Resilience
Mission.

**** *auteur_EM *date_24_03_2022 *spécificité_EM2
La résolution humanitaire portée par la France et le Mexique a été votée à une très large
majorité à l'Assemblée générale des Nations unies : 140 voix pour, 5 voix contre. Cela
montre l'isolement que nous maintenons sur la Russie.

**** *auteur_EM *date_24_03_2022 *spécificité_EM2
Avec nos Alliés, nous allons continuer à fournir des armements, dans un cadre d'efficacité
complète, avec une ligne rouge : ne pas être co-belligérants.
```

Figure 11
Balisage du corpus.

Sur la base de ce découpage, le chercheur peut comparer les spécificité (sur- et sous-emploi) en choisissant des termes qui l'intéressent. Nous avons choisi d'analyser des mots en lien avec l'éducation et avec l'écologie.

Concernant l'éducation, nous avons sélectionné les lemmes « élève », « éducation », « école », « enseignement » et « enseignant » et procédé à une représentation graphique de leur sur- ou sous-emploi dans les parties de corpus (Figure 12) :

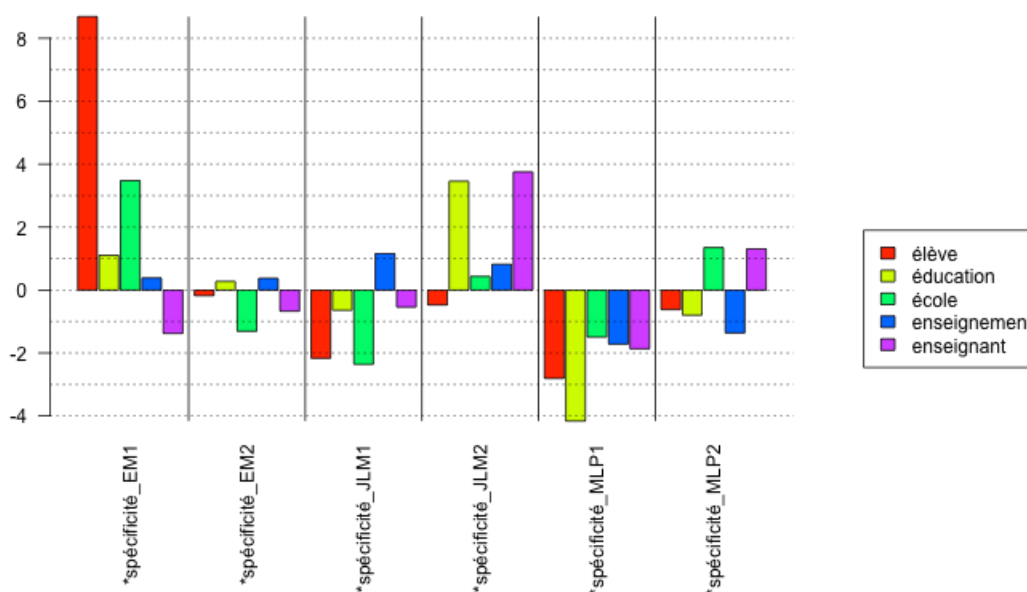


Figure 12
Spécificités de termes liés à l'éducation.

On note que les candidats qui sur-utilisent ces termes sont Emmanuel Macron 2017 (« élève » et « école » notamment ainsi que « éducation », Jean-Luc

Mélenchon 2022 (avec « éducation » et « enseignant » en particulier), et Marine Le Pen 2022 dans une moindre mesure (avec « école » et « enseignant »). Il faut bien sûr ensuite retourner au corpus pour mener l'analyse complète (qui n'est pas le propos de cet article) mais on voit bien comment l'accès à ce corpus réalisé en « temps réel » lors de campagnes peut ensuite donner lieu à des analyses sociodiscursives plus complexes. La même chose peut être réalisée autour des mots de l'écologie (Figure 13), que l'on peut utiliser pour montrer la prégnance de « écologique » chez Jean-Luc Mélenchon (et son sous-emploi chez Emmanuel Macron 2017 et Marine Le Pen), l'emploi de « environnemental » et « environnement » chez Emmanuel Macron en 2017, remplacé par « climatique » en 2022 :

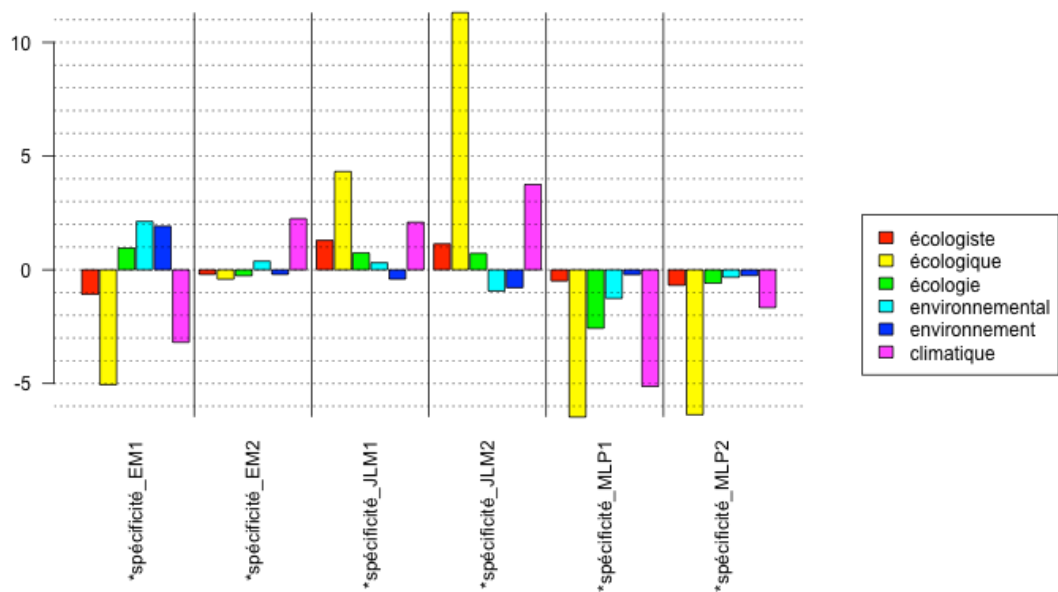


Figure 13
Spécificités de termes liés à l'écologie.

L'analyse des spécificités a donc l'avantage de pouvoir « mesurer » les corpus non pas à travers le calcul des fréquences, mais en lien avec l'importance du vocabulaire de chaque partie relativement à l'ensemble du corpus.

Cette analyse comparative peut être étendue, en termes de variables et de perspectives, grâce au recours au *deep learning*.

4.2. Le deep learning et l'adaptation à plusieurs niveaux d'analyse

Comme nous l'avons expliqué à propos de « Mesure du discours », le recours au *deep learning* permet un nouveau type de traitement, qui permet notamment de classifier les résultats en fonction de certaines problématiques. Dans un travail mené à propos de l'attribution d'auteur (Lam, Demange,

Longhi 2021), le paramétrage suivant (Figure 14) a été établi, pour donner les résultats optimums sur l'analyse d'un corpus de tweets :

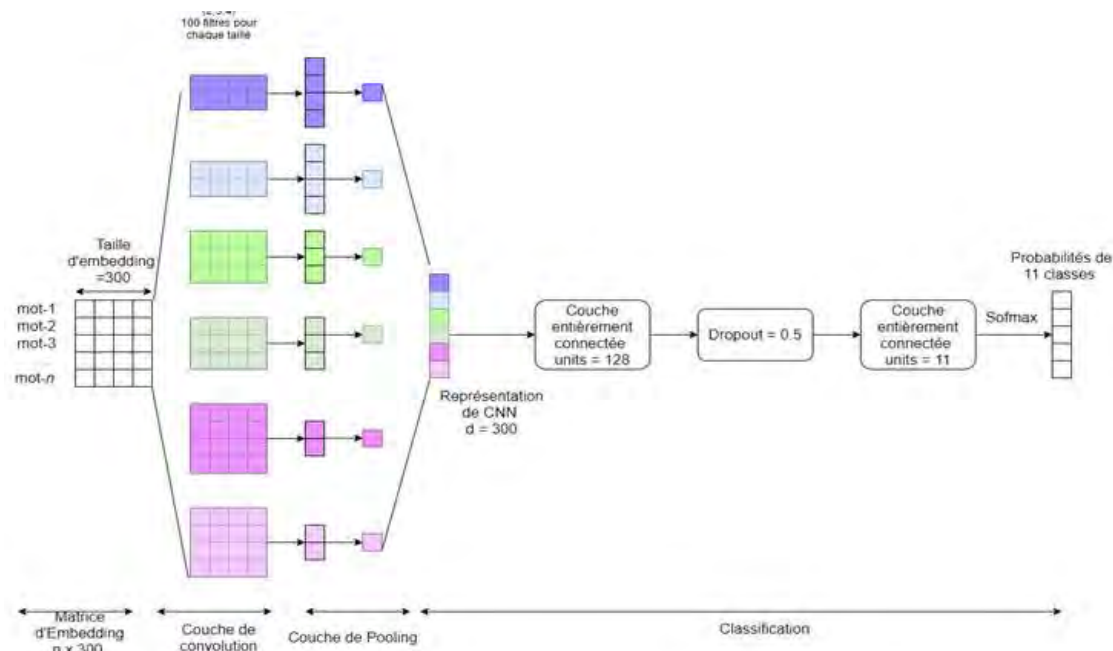


Figure 14
Choix du modèle et paramétrage du classifieur.

Ce modèle a été appliqué au corpus précédemment étudié avec Iramuteq, afin d'observer si le classifieur prédit correctement l'auteur du message. Nous représentons les résultats dans la matrice de confusion en Figure 15, qui confronte les classes prédites (en ordonnées) aux vraies classes (en abscisse) :

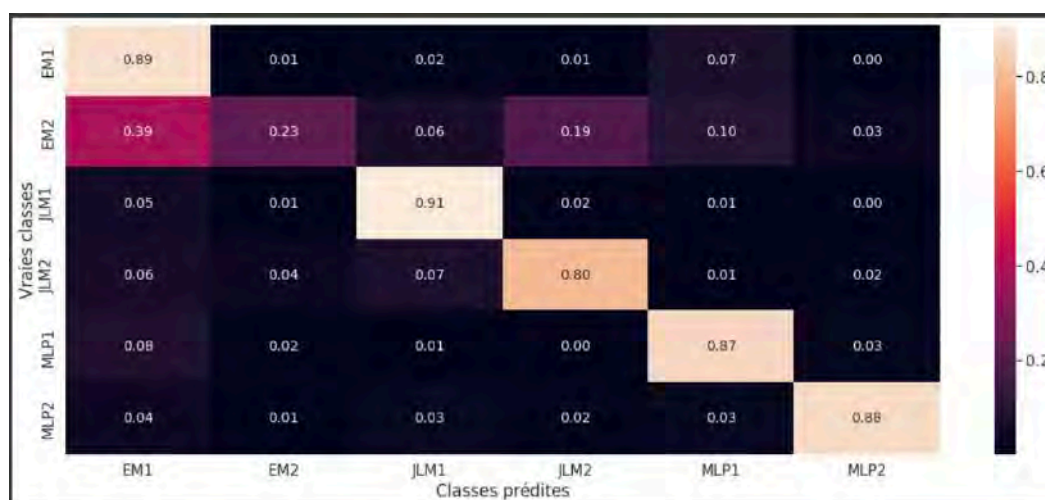


Figure 15
Matrice de confusion (classes prédites/vraies classes) avec les 6 variables.

On observe ainsi que la plupart des résultats sont corrects voire bons (diagonale de la figure), avec par exemple un score de 0,91 de bonne

prédiction (précision) pour Jean-Luc Mélenchon 2017, ou 0,89 pour Marine Le Pen 2022. Mais la ligne EM2 (Emmanuel Macron 2022) montre une disparité forte avec le reste des résultats, puisque le score est de seulement 0,23 pour les bonnes prédictions (alors que le classifieur prédit 0,39 pour EM1, et 0,19 avec Jean-Luc Mélenchon 2022). À ce premier niveau, les « confusions » du classifieur sont intéressantes, puisque cela nous permet d'anticiper sur d'éventuelles proximités ou similarités entre candidats.

Comme préconisé par Vanni et ses collègues, le principe de déconvolution est ici utile pour rendre ces classifications intelligibles. Dans les deux figures suivantes, on observe des erreurs de classification.

Dans l'exemple suivant (Figure 16), le classifieur indique EM1 (Emmanuel Macron 2017) alors que le vrai label est MLP1 (Marine Le Pen 2017) :

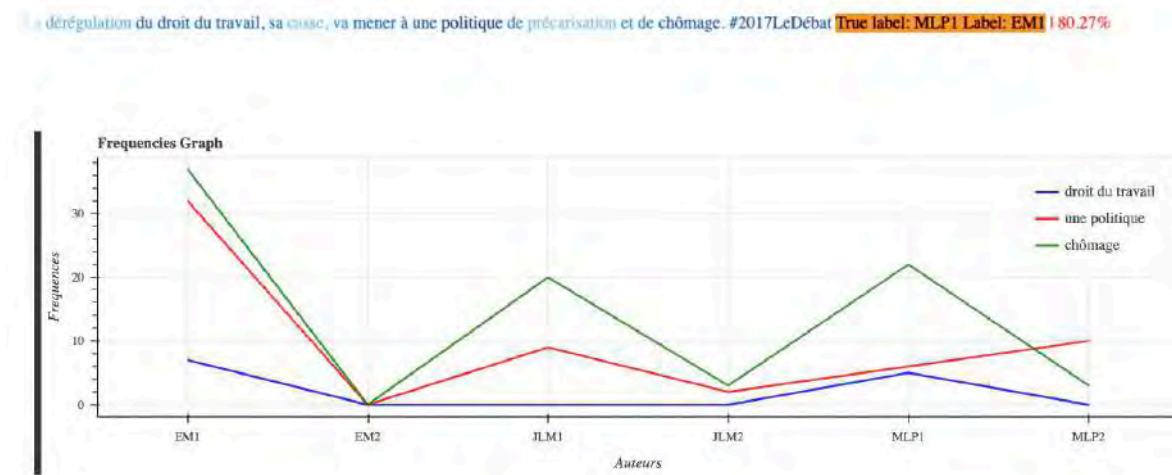


Figure 16

Déconvolution à propos d'un exemple de MLP1 attribué à EM1.

Les expressions/termes « droit du travail », « une politique » et « chômage » sont distingués comme caractéristiques du discours de d'Emmanuel Macron en 2017. Or dans ce message de Marine Le Pen, on a la coïncidence de ces trois marqueurs : cela explique la confusion du réseau, qui a attribué le message à EM1.

Dans le second exemple (Figure 17), le classifieur fait cette fois une confusion entre EM1 et JLM1 : à partir de la co-présence de « tribune », « un moment de », « débat » et « convaincre », les classifieur penche pour EM1 alors qu'il s'agit de JLM1 :

Une élection n'est pas pour moi une simple tribune : c'est un moment de débat où il faut convaincre. #RTLMatin True label: JLM1 Label: EM1 | 99.14% |

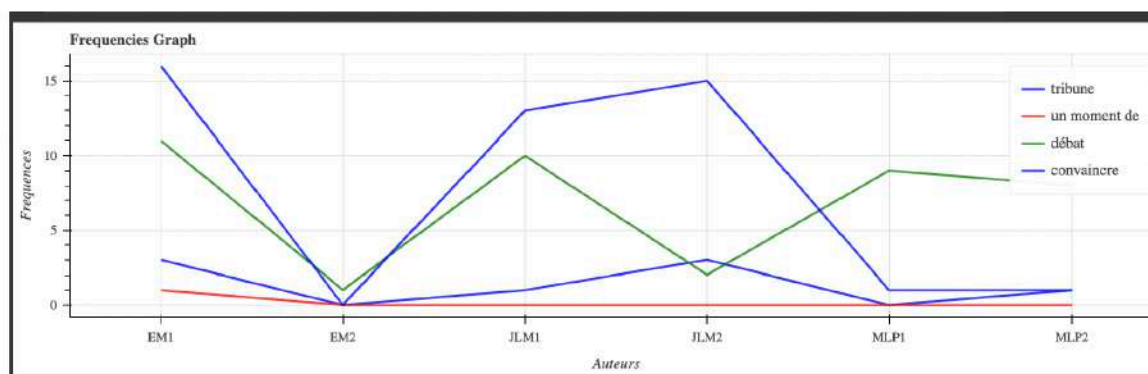


Figure 17

Déconvolution à propos d'un exemple de JLM1 attribué à EM1.

Ces confusions sont utiles pour l'analyste du discours intéressé par le discours politique, car elles révèlent des enjeux politico-discursifs importants : proximité entre candidats, adaptation des discours aux contextes, proximités lexicales sur certains thèmes, etc. Dans le cas qui nous intéresse, une particularité autour du candidat EM2 tient au fait que dans la campagne 2022, il est le président sortant, et a donc un comportement discursif très différent. Aussi, pour compléter l'analyse, nous avons cherché les comptes qui pourraient se faire le relai du « candidat Macron », puisque le compte officiel coïncide avec celui du président (ce qui entretiendrait une certaine confusion des genres).

4.3. Prise en compte du contexte politique et élargissement du corpus

Nous avons ainsi repéré deux comptes susceptibles de « faire campagne » en lieu et place de celui du président : « En Marche » (EMA), et « avec vous », caractérisé par « Emmanuel Macron avec vous ! » (Figure 18) :



Figure 18
Description des 2 comptes ajoutés pour la campagne 2022.

Pour savoir si cela a une influence sur les résultats, et l'analyse par le modèle, on peut lancer une nouvelle analyse et observer dans la Figure 19 les prédictions du modèle :

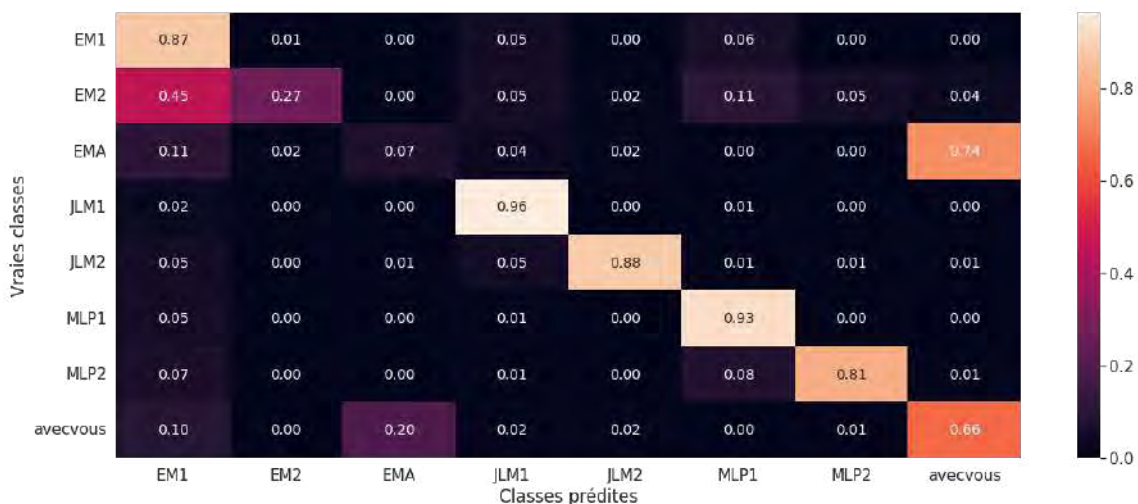


Figure 19
Matrice de confusion (classes prédites/vraies classes) avec les 8 variables.

La ligne EMA est particulièrement intéressante, puisque les prédictions sont extrêmement mauvaises (0,07), et les erreurs les plus importantes sont avec EM1 (0,11) et surtout avec avecvous (0,74). Les score pour avecvous sont très moyens (0,66), et les erreurs sont essentiellement avec EMA (0,20) et EM1 (0,10). Les erreurs ont donc un sens, puisqu'il y a des confusions très fortes, dans les différentes configurations, entre EMA, avecvous, et EM1. Le style d'argumentation électoral macronien se transpose donc ailleurs en 2022 pour ne pas se confondre avec la parole présidentielle, et des relais sont mis

en place pour conserver une continuité avec la posture argumentative du contexte de la campagne.

5. Conclusion

L'évolution des technologies numériques a profondément transformé la manière dont nous échangeons des idées politiques. La montée en puissance des réseaux sociaux, des blogs et des forums de discussion en ligne a offert une plate-forme de plus en plus accessible pour les organisations politiques de communiquer et d'interagir avec les électeurs. L'une des principales évolutions du discours politique dans le contexte numérique est l'accent mis sur l'interaction et la participation. Les candidats et partis politiques utilisent les réseaux sociaux pour atteindre un public plus large et pour mobiliser leurs sympathisants lors des élections. Cette évolution du discours politique n'est pas sans ses défis : les plateformes de médias sociaux ont souvent été critiquées pour favoriser la polarisation et l'incitation à la haine. Aussi, l'analyse du discours politique numérique est devenue de plus en plus importante en raison de l'importance croissante d'Internet et des médias sociaux dans le débat public et la formation de l'opinion. Les nouveaux défis de l'analyse du discours politique numérique concernent notamment la quantité de données (qui sont souvent bruyantes et contiennent des informations redondantes ou peu pertinentes), la polarisation, la désinformation, et la difficile distinction entre les faits, les opinions et les rumeurs, notamment dans des messages courts. L'analyse du discours politique numérique est un domaine en constante évolution, avec de nouveaux défis à relever : les chercheurs doivent être capables d'adapter leurs méthodes et leurs outils pour répondre à ces défis et produire des analyses précises et pertinentes.

Dans cet article, nous avons présenté différents projets qui, au cours des élections présidentielles françaises 2017 et 2022, ont contribué à rendre accessibles les méthodes, outils et résultats de la recherche auprès d'un plus grand nombre. En détaillant notamment certains enjeux scientifiques autour de la classification des candidats, nous avons pu montrer l'intérêt scientifique et citoyen de tels projets, qui renouvellent tout autant la manière de faire de l'analyse du discours, que l'analyse des observables eux-mêmes, et les critères mis en œuvre pour faire émerger des résultats. La transformation numérique est donc un moyen de renouveler les terrains, méthodes, et modes de diffusion, de l'analyse du discours, et devient une opportunité pour observer les évolutions et mutations de la communication politique.

Bionote: Julien Longhi is *Professeur des universités* at CY Cergy Paris Université. He holds a PhD in discourse analysis at the University of Clermont-Ferrand II, and a *Habilitation à diriger des recherches* at CY Cergy Paris Université. His research focuses on the question of ideologies in political discourse, and on the use of discourse analysis to explore corpora. He is the director of the Digital Humanities Institute at CY, and has published several books and articles in the fields of digital discourse analysis, digital humanities, and toolled linguistics.

Author's address: julien.longhi@cyu.fr

Acknowledgements: We would like to thank CY Fondation and the National Research Agency (ANR) for their precious financial support to the current research.

Références bibliographiques

- Brunet E., Lebart L. et Vanni L. 2021, *Littérature et intelligence artificielle*, in Mayaffre D. et Vanni L. (éds.), *L'intelligence artificielle des textes*, Honoré Champion, Paris, pp. 73-130, coll. Lettres Numériques.
- Gaumont G., Panahi M. et Chavalarias D. 2018, *Reconstruction of the socio-semantic dynamics of political activist Twitter networks -Method and application to the 2017 French presidential election*, in "PLOS ONE", <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201879>.
- Lam T., Demange J. et Longhi J. 2021, *Attribution d'auteur par utilisation des méthodes d'apprentissage profond*, in « EGC 2021 Atelier DL for NLP : Deep Learning pour le traitement automatique des langues ». <https://hal.science/hal-03121305/document>
- Lebart L. et Salem A. 1994, *Statistique textuelle*, Dunod, Paris.
- Longhi J. 2008, *Objets discursifs et doxa : essai de sémantique discursive*, L'Harmattan, Paris.
- Longhi J. 2015, *L'acte de nommer comme constitution de formes : discursivité de la production du sens*, in « Langue française » 188 [4], pp. 121-136.
- Longhi J. 2016, *Le tweet politique efficace comme même textuel : du profilage à viralité*. in « Travaux de linguistique » 73 [2], pp. 107-126.
- Longhi J. 2019, *Le projet #Idéo2017 : Quelles implications du/de la chercheur-e en tant qu'acteur-trice potentiel du changement social ? Exemplification à partir du discours politique numérique*, in « Cahiers de Linguistique » 44 [2], pp. 103-116.
- Longhi J. 2020, *Proposals for a discourse analysis practice integrated into digital humanities: theoretical issues, practical applications, and methodological consequences*, in "Languages" 5 [1], 5. <https://doi.org/10.3390/languages5010005>
- Mayaffre D. et Vanni L. (éds.) 2021, *L'intelligence artificielle des textes. Des algorithmes à l'interprétation*, Honoré Champion, Paris.
- Mikolov T., Sutskever I., Chen K., Corrado G.S. and Dean J. 2013, *Distributed Representations of Words and Phrases and Their Compositionality*, in Burges C., Bottou L., Welling M., Ghahramani Z. and Weinberger K. (eds), *Advances in Neural Information Processing Systems*, Vol. 26, Curran Associates, Inc., Red Hook pp. 3111–3119.
- Pincemin B. 2012, *Hétérogénéité des corpus et textométrie*, in « Langages » 187, pp. 13-26. <https://doi.org/10.3917/lang.187.0013>.
- Vanni, L. et Precioso, F. (2021). *Deep learning et description des textes. Architecture méthodologique*, in Mayaffre D. et Vanni L. (éds.), *L'intelligence artificielle des textes. Des algorithmes à l'interprétation*, Honoré Champion, Paris, pp. 73-130, coll. Lettres Numériques.

DISCOURS NUMERIQUES ET APPROCHES OUTILLÉES

Quelques réflexions sur les apports des métadonnées

CLAUDIA CAGNINELLI
UNIVERSITÀ DEGLI STUDI DI MILANO

Abstract – This paper examines the role of metadata and annotations in the analysis of digital corpora by linguistic analysis tools. The present study is grounded in a discursive approach to the analysis of linguistic data. This conceptual framework determines both the concept of the corpus adopted for the study and the methodological principles guiding its exploration. The structuring of the corpus through metadata represents a pivotal juncture between the digital constitution of the material observed and the way in which it can be explored by means of corpus analysis tools. The paper initially addresses the methodological transformations, and the epistemological implications produced by digital resources. The evolving conceptualization of empirical data and the nature of the objects of study are also highlighted. A typology of metadata is then proposed, based on two main parameters: the different types of information they represent, on the one hand, and the characteristics of the data, on the other. Specific focus is directed towards the digital discourses of Web 2.0.

Keywords: corpus; digital discourse; corpus analysis tools; corpus structure; annotations.

1. Introduction

Les évolutions technologiques qui accompagnent nos sociétés de même que la recherche scientifique, y compris les études en sciences humaines et sociales, ont transformé les données soumises à l'observation, leurs formats, les objets d'étude ainsi que les ressources et les approches méthodologiques adoptées pour les analyser. En particulier, les corpus en format numérique représentent désormais un outil fondamental et de plus en plus diffus grâce à une accessibilité facilitée aux ressources numériques pour la constitution et l'exploration des corpus (cf. Condamines 2022). Comme le remarquait Mayaffre en 2010, le corpus a gagné un rôle central et incontournable dans les études linguistiques dès le début du nouveau millénaire, au point que « la linguistique *sans* ou *hors corpus* paraît aujourd'hui une spéculation intellectuelle marginale pratiquée seulement par une minorité » (Mayaffre 2010, p. 233).

Si les corpus numériques et le recours à des analyses outillées sont de

plus en plus nécessaires à divers domaines des recherches linguistiques, il faut toutefois bien souligner que la conception, la constitution et les démarches d'analyse du corpus changent selon les cadres théoriques, les objets et les objectifs d'étude ainsi qu'en fonction des méthodes d'exploration adoptées (Condamines 2005 ; Mayaffre 2005 ; Cori *et al.* 2008). Il nous semble ainsi fructueux d'interroger les fondements théorico-méthodologiques qui caractérisent les évolutions de la conception du corpus et de ses principes d'exploration au sein du domaine où se situe la recherche du linguiste, en opérant aussi un retour réflexif à partir d'expériences personnelles d'application.

S'inscrivant dans une approche discursive qui repose sur une « conception *philologique-herméneutique* » du corpus (Rastier 2011, p. 35), notre contribution se propose d'aborder dans une perspective théorico-méthodologique la question des métadonnées dans la structuration des corpus numériques. Nous traiterons notamment leur portée dans la création des conditions nécessaires pour l'appréhension d'observables discursifs dans le cadre d'une analyse outillée. Nos réflexions prennent appui sur la recherche menée pour notre thèse (Cagninelli 2022), ainsi que sur d'autres études que nous avons réalisées sur des corpus numériques relevant de plusieurs types et genres de discours, et les prolongent. Tout d'abord, les transformations méthodologiques provoquées par l'introduction de l'informatique dans la constitution et l'exploration de corpus en linguistique seront mises en relation avec les enjeux épistémologiques qu'elles ont entraînés sur l'appréhension des données linguistiques et des objets d'étude (Section 2). Nous aborderons ensuite le rôle des métadonnées dans la structuration des corpus numériques et nous proposerons une typologie de métadonnées sur la base des informations dont elles rendent compte et de la nature des données qu'elles renseignent (Section 3). Après une bipartition principale entre métadonnées contextuelles et analytiques (3.1), nous approfondirons le cas des métadonnées de décomplexification (3.2), qui s'avèrent nécessaires pour l'étude outillée de textualités numériques du web 2.0 à la lumière de leurs spécificités distinctives. La réflexion théorico-méthodologique et l'opérationnalité de la classification proposée seront enfin illustrées dans la section 4 à travers une exemplification aussi bien de la structuration d'un corpus numérique que de ses possibles applications à des fins d'analyse.

2. Corpus numériques et explorations outillées : transformations méthodologiques et implications épistémologiques

L'informatique a marqué un tournant dans l'évolution du format et de la conception même des corpus, tout comme dans les méthodes et les outils

destinés à leur exploration. De même, elle a eu des répercussions tant sur la nature des données à soumettre à l'observation – envisagées comme les matériaux langagiers sélectionnés et abordés dans une perspective linguistique – que sur celle des objets d'étude d'un point de vue à la fois théorique et méthodologique.

Le passage au format numérique a influencé aussi bien la visualisation des données en corpus que les modes d'accès aux données. La quantité de données à examiner a augmenté notablement et les fonctionnalités de recherche intégrées aux environnements numériques en ont facilité la consultation. Ce sont toutefois les calculs statistiques et les méthodes d'analyse intégrées aux outils d'exploration exploités par la lexicométrie et plus tard par la textométrie et la logométrie qui ont jeté les bases de l'« herméneutique numérique » (Mayaffre 2002). Comme le fait remarquer Mayaffre :

Par une lecture différente (hypertextuelle plutôt que linéaire, nous l'avons vu, mais aussi paradigmatique plutôt que syntagmatique, quantitative plutôt que qualitative), l'ordinateur voit autre chose pour déranger nos certitudes et élargir l'horizon étroit de nos *modes* (aux deux sens du terme) d'interrogation. (Mayaffre 2002, p. 155)

De surcroît, le format numérique des corpus présente l'avantage de pouvoir enrichir la matérialité linguistique des productions empiriques sélectionnées, en ajoutant toutes sortes d'informations pertinentes pour leur étude sous forme de métadonnées. Les métadonnées sont des informations que l'on peut associer aux matériaux langagiers et leur mise en corpus contribue à ce qu'elles soient appréhendées en tant que données linguistico-discursives à partir du cadre théorico-méthodologique sur lequel repose la conceptualisation et la structuration du corpus.

Prenons l'exemple d'un corpus textuel créé au format XML. Avec ce type de format, il est possible d'insérer les données linguistiques au sein d'une structure arborescente constituée de balises qui permettent d'ajouter des informations à différents niveaux. Le tableau 1 illustre comment ces informations peuvent être ajoutées sous forme de métadonnées en relation avec l'unité textuelle de base envisagée pour la constitution du corpus. La colonne gauche restitue la partie verbo-textuelle d'un discours numérique de nature « technodiscursive » (Paveau 2017) – il s'agit en l'occurrence d'un tweet¹ publié par le compte @EmmanuelMacron le 12 décembre 2020 (cf.

¹ Le tweet représente le format de base des publications via le réseau socio-numérique Twitter, rebaptisé X en juillet 2023. Considérant que les exemples proposés datent de 2020, nous adopterons ici les dénominations répandues dans la littérature académique en vigueur jusqu'au changement de nom de la plateforme. Nous précisons également que plusieurs des

infra Figure 1) –, alors que la colonne droite propose un exemple de sa mise en forme pour l'intégrer dans un corpus au format XML. Dans ce cas, la partie verbo-textuelle du tweet est structurée en relation avec des informations concernant l'identification de l'unité textuelle dans le corpus (balise *texte*), la source énonciative (*locuteur*) ainsi que la date de production (*date*).

Données linguistiques du corpus retenues à partir de la matérialité langagière de la production sélectionnée	Données enrichies de métadonnées au format XML
Avons-nous atteint tous nos objectifs #5AnsAprès l'Accord de Paris sur le climat ? Non. En revanche, ne sous-estimons pas le chemin parcouru. Je m'étais engagé à rendre des comptes : c'est ce que je veux faire aujourd'hui.	<pre> <texte="1"> <locuteur="EMacron"> <date="12/12/2020">Avons-nous atteint tous nos objectifs #5AnsAprès l'Accord de Paris sur le climat ? Non. En revanche, ne sous-estimons pas le chemin parcouru. Je m'étais engagé à rendre des comptes : c'est ce que je veux faire aujourd'hui.</date> </locuteur> </texte> </pre>

Tableau 1

Exemple de mise en forme d'une unité textuelle du corpus au format XML avec l'ajout de métadonnées qui définissent une structuration possible au niveau textuel.

Si la balise *texte* sert ici à reconnaître de manière univoque le tweet en tant qu'unité textuelle spécifique du corpus, les balises *locuteur* et *date* apportent des éléments d'information relevant du contexte de production du discours. Ces métadonnées peuvent ainsi être exploitées pour opérer différents types de regroupements et de comparaisons des données au service de plusieurs perspectives d'observation, à l'instar d'études qui prennent en compte la variation au fil du temps ou qui comparent les discours de locuteurs différents.

Les balises envisagées à titre d'exemple dans le tableau 1 définissent la structure d'une unité du corpus textuel, tout en façonnant en même temps la structure de base du corpus. Par ailleurs, ce dernier pourrait également prévoir des métadonnées supplémentaires aussi bien au niveau macro-textuel (à l'instar de la classification de plusieurs genres textuels – cf. section 3.1) qu'aux niveaux micro- et mésotextuels, comme c'est le cas pour l'ajout d'annotations d'ordre morpho-lexical, sémantique ou syntaxique, qui ne seront toutefois pas traitées dans la présente étude.

fonctionnalités de la plateforme ont été progressivement renommées aussi à partir de juillet 2023.

Comme nous le verrons plus loin (Section 3), il est en effet possible de distinguer plusieurs macro-catégories de métadonnées pour *informer* – au sens de modeler, de donner une forme spécifique – la matérialité langagière des productions socio-discursives. Cependant, le type et la quantité de métadonnées à prévoir dépendent aussi bien de la nature du discours sélectionné que des objectifs de l'étude et des méthodes d'analyse adoptées².

La constitution d'un corpus en format numérique et enrichi de métadonnées permet de multiplier non seulement les informations associées aux données, mais aussi les perspectives d'appréhension et les paramètres à prendre compte au cours de l'analyse.

Par ailleurs, les données rassemblées en corpus sont susceptibles de se diversifier, tant du point de vue de la nature de leurs constituants sémiotiques que des textualités dont elles relèvent. En effet, les technologies numériques ont également contribué à la naissance de pratiques communicationnelles nouvelles et en conséquence de données et notamment de formes de textualités tout aussi nouvelles. On peut mentionner à titre d'illustration le cas des « technodiscours » (Paveau 2017), comprenant les productions issues du web 2.0, comme les posts des réseaux socio-numériques, dont la nature est à la fois langagière et technologique, ainsi que le remarque Paveau. Ces textualités créées par le biais de dispositifs connectés présentent des traits distinctifs inédits qui appellent la recherche linguistique à questionner ses cadres épistémologiques et méthodologiques pour les étudier, à l'instar de la proposition avancée par Paveau (2017). Les technodiscours théorisés par Paveau (2017) se caractérisent en effet par un potentiel de *délinéarisation* qui peut ouvrir divers parcours de lecture grâce à la présence de « formes technolangagières », telles que les hashtags ou les liens URL, participant d'ailleurs de l'*augmentation énonciative* de ces productions. L'auteure souligne également leur nature *composite* et *relationnelle*, ainsi que l'*investigabilité* et l'*imprévisibilité* qui est propre à leur matérialité numérique.

Cependant, il est évident que les données qui ne sont pas natives – au sens de Paveau (2017) – d'un environnement numérique peuvent à leur tour donner lieu à des corpus en format numérique. Paveau (2015, en ligne, §13) distingue en effet « trois ordres linguistiques du numérique » en fonction de plusieurs critères tels que « le mode de production de l'écrit, le mode de lecture du texte et son inscription dans l'écosystème du Web ». Aux deux extrêmes de la tripartition proposée par Paveau (2015), les textualités numériques non natives ou « numérisées » s'opposent aux textualités

² Il faut également noter que le format du fichier à analyser doit être compatible avec les formats utilisés par le logiciel d'analyse. Cela a donc des répercussions sur la mise en forme des données, sans pour autant affecter nécessairement leur conceptualisation.

numériques natives ou « numérisées » correspondant aux technodiscours du web 2.0. Puisque les textualités numérisées résultent de textes qui sont importés en contexte numérique par des opérations de numérisation, elles ne présentent pas les traits définitoires des technodiscours cités plus haut. En l'occurrence, un processus de médiation est par conséquent nécessaire, comme c'est le cas d'ailleurs pour tout passage du médium d'origine à un autre support, à l'instar de la transcription écrite d'une production orale.

Cela ne signifie toutefois pas que les discours numériques natifs ne soient pas soumis à des opérations d'interprétation lors de la constitution du corpus. Il faut en effet rappeler que le corpus est un objet construit (cf. entre autres, Dalbera 2002 ; Mayaffre 2005 ; Garric, Longhi 2012), qui rassemble des matériaux linguistiques empiriques selon des critères de sélection établis par l'analyste en fonction d'un cadre théorico-méthodologique et d'un objet d'étude déterminés. De fait, tout groupement de textes ne correspond pas nécessairement à un corpus, selon l'acception que ce dernier assume en sciences du langage, comme le montre Duteil-Mougel (2006 ; cf. aussi Rastier 2011). Le corpus dépend des conditions de sa constitution (cf. Rastier 2011 ; Pincemin 1999, 2012). En particulier, il doit respecter un « critère d'interprétabilité » (Pincemin 2012) afin de matérialiser un « terrain » représentatif pour l'objet envisagé (Longhi 2018).

En bref, les transformations engendrées par les technologies, les dispositifs et les outils informatiques ont affecté tant les formes de la matérialité langagière – participant de différents ordres textuels (cf. *supra*) – que les modalités et les moyens de son observation. Ces transformations ont ainsi offert à l'attention du linguiste de nouvelles textualités technodiscursives – façonnées par le numérique et composées du technologique – et de nouveaux objets d'étude, dont l'appréhension est rendue possible par divers outils et méthodes d'exploration informatisée. Plusieurs chercheurs (Mayaffre 2010 ; Rastier 2011 ; Neveu 2016) soulignent les changements épistémologiques profonds entraînés par un « rapport renouvelé à l'empirique » (Rastier 2011, p. 47 ; Neveu 2016, p. 3), rendu également possible par la constitution de corpus numériques³ et par le développement d'outils d'analyse informatiques. Comme l'affirme Rastier,

Un nouveau rapport à l'empirique change non seulement l'étendue mais la nature des faits et rend nécessaire l'innovation théorique. Il permet de produire de nouveaux faits, qui naissent pour ainsi dire de la rencontre entre de nouveaux modes d'observation et d'explication. (Rastier 2011, p. 47)

³ Par *corpus numériques*, nous désignons les corpus en format numérique, qu'ils soient constitués ou non de textes nativement numériques.

Mayaffre (2010, p. 240) soutient en outre qu'« [a]vec les grands corpus numériques, c'est à la fois notre rapport à l'empirie du langage qui est modifié et une nouvelle heuristique des phénomènes langagiers ou discursifs qui peut être envisagée ».

Les apports des corpus numériques et des explorations outillées concernent donc plusieurs plans interreliés : (1) la constitution des ressources à observer – ce qui a des répercussions sur la taille et le format des données, le cadrage et les catégorisations théoriques associées ; (2) l'approche des données, transformée par le formatage numérique et l'exploration outillée ; (3) l'émergence des observables ; (4) les apports théoriques de ces nouveaux objets d'observation, méthodes d'observation et observables, et finalement leurs implications épistémologiques pour les études linguistiques. La mise en corpus numérique – c'est-à-dire la structuration en format électronique du matériau linguistique attesté par des réalisations authentiques de l'activité langagière – permet d'obtenir des données linguistiques empiriques analysables en fonction du cadre théorico-méthodologique de l'analyste ainsi que des objectifs de son analyse. Cette mise en corpus constitue par conséquent une étape fondamentale pour les appréhender à partir d'un cadrage et de catégorisations théoriques spécifiques, matérialisées par la structure du corpus.

La structuration du corpus peut en effet avoir des répercussions sur la conceptualisation des observables ainsi que sur leur identification. Sans approfondir la question des observables dans le cadre de cette étude, nous précisons pourtant l'acception dans laquelle nous utiliserons ce mot qui est susceptible d'emplois variés dans la recherche en sciences du langage, ainsi que le signale Neveu (2016). Suivant les travaux de Neveu (2016) et de Constantin de Chanay et Ferron (2017), nous envisageons l'observable comme ce qui, à partir de la matérialité attestée et observée en corpus, est interprété par l'analyste comme étant pertinent et relevant de l'objet d'étude théorisé. Notre utilisation du terme *observable(s)* vise donc à attirer l'attention sur le lien de pertinence, établi par l'analyse interprétative du linguiste, entre la matérialité attestée en corpus et l'objet d'étude ancré dans un cadre théorique spécifique. Ce lien résulte de la relation entre les dimensions empirique et théorique qui est présupposée par le concept d'*observable(s)* tel que l'envisagent à la fois Neveu (2016), en s'appuyant aussi sur le travail de Culioli (1968), et Constantin de Chanay et Ferron (2017). Ces derniers mettent en outre en évidence la dimension même virtuelle des observables, qui peuvent s'actualiser par l'analyse d'un corpus, lorsqu'ils remarquent l'interrelation étroite des activités théorique et pratique de l'analyse linguistique :

Parler d'observable implique donc deux choses : non seulement que la linguistique soit considérée comme une science empirique, mais aussi que

l'observé soit théorisé : qu'il soit défini, qu'il soit extensible à du non encore observé, que l'on puisse établir un lien décisif entre lui et l'objet scientifique que l'on poursuit à travers lui. (Constantin de Chanay, Ferron 2017, p. 8)

L'analyse du corpus donne ainsi accès à la matérialité linguistique à partir d'une perspective déterminée, celle qui est établie par l'analyste en fonction de son objet d'étude. Par l'exploration du corpus, l'analyste met en relation l'empirique avec le théorique, en repérant les traces, les faits et les phénomènes qui caractérisent et déterminent l'objet d'étude. Il s'ensuit que les unités d'observation changent selon l'objet d'étude et que la multiplication des angles d'observation enrichit non seulement l'appréhension des données mais aussi la possibilité d'observables pertinents. La variété de perspectives d'analyse matérialisées par la structure du corpus permet en effet d'avoir accès et de vérifier la théorisation d'objets d'étude plus complexes.

La façon dont la matérialité langagière est structurée en corpus permet donc de l'appréhender différemment de celle d'une observation hors corpus. En combinant plusieurs méthodes d'exploration, on a en outre la possibilité d'étudier les mêmes données à travers des découpages différents et des angles d'approche variés concrétisés par la structure du corpus. Comme nous l'illustrerons dans la section 3, les différents types de métadonnées envisagées peuvent apporter aussi bien des informations factuelles, relatives notamment au contexte des productions ainsi qu'à leur format structurel, que des catégorisations et donc des conceptualisations théorico-méthodologiques des matériaux sélectionnés. Selon les objets et les objectifs de l'étude, les métadonnées d'ordre théorico-méthodologiques donnent corps à plusieurs niveaux d'observation de la textualité – à travers l'articulation du local et du global (cf. Rastier 2001, 2011 ; Mayaffre 2005) – tout comme à différents niveaux possibles d'analyse linguistique et (techno)discursive (cf. section 3.2). L'approche des données devient ainsi multidimensionnelle, plus riche quantitativement et qualitativement, et finalement plus complexe.

3. Structuration des corpus numériques : types et rôles des métadonnées

L'activité de structuration d'un corpus détermine son organisation interne. Pour les corpus en format numérique, cette opération se réalise par l'ajout d'informations et de descripteurs aux données, selon différentes unités d'observation et divers niveaux de découpage et regroupement. Par la mise en corpus, la matérialité langagière sélectionnée est structurée et balisée en fonction des exigences méthodologiques imposées par sa nature, de même que par l'objet d'étude envisagé. Ainsi, si les données du corpus constituent les sources primaires qui font l'objet de l'observation, les métadonnées

structurant le corpus pourraient être considérées comme des sources secondaires qui les enrichissent (cf. section 3.1), voire les complètent dans le cas de textualités technodiscursives et plurisémiotiques (cf. section 3.2).

Par conséquent, les métadonnées informent et orientent l'observation. Elles jouent un rôle clé (1) pour restituer des informations du contexte de production qui sont significatives aux fins de l'étude, (2) pour rendre compte et identifier des éléments constitutifs des textualités⁴ sélectionnées, ainsi que (3) pour projeter des catégorisations théoriques sur les données. Le format numérique du corpus et notamment sa structuration par le biais de différents types de métadonnées contribuent à concrétiser une pluralité de formes de contextualisation et d'enrichissement des données. Ces informations associées aux données peuvent être de nature aussi bien factuelle, lorsqu'elles concernent en particulier le contexte de production du discours et son format structurel, que théoriques, lorsqu'elles établissent des catégorisations des données à partir du cadre théorique et de l'approche méthodologique adoptés. La « corporalité » du corpus (Mayaffre 2010) et ses matérialisations au cours de l'analyse s'avèrent donc incontournables pour l'actualisation⁵ des observables. Le corpus représente ainsi

le lieu de réalisation, et le niveau d'accès, à/(de) la prise sociocognitive et langagière du monde par le sujet, dont la matérialité langagière corporéifie la saisie. Le corpus rend compréhensible cette dynamique du niveau le plus local (palier du morphème ou mot) à l'organisation textuelle, jusqu'à l'insertion en discours, selon une pratique sociale et une orientation argumentative donnée. (Longhi, Sarfati 2018, p. 105)

La structuration du corpus contribue de manière significative à sa corporalité, en raison des différentes configurations des données qu'il est susceptible de matérialiser et d'offrir au regard de l'analyste lors de son exploration outillée. Comme nous le montrerons dans la section 3.2, la prise en compte d'éléments intégrés au corpus en tant que métadonnées permet par exemple d'analyser les caractéristiques de la partie verbo-textuelle des technodiscours en fonction de la présence de différents contenus pluritechnosémiotiques qui participent à la création du sens global du message – des images, aux vidéos, aux GIFs ou encore aux mèmes. Différentes similitudes et oppositions peuvent également ressortir des catégorisations théoriques projetées sur les données ainsi que des paramètres et des niveaux d'observation adoptés, qui peuvent se multiplier au cours de l'analyse.

⁴ Précisons que notre réflexion porte sur des corpus rassemblant une ou plusieurs formes de textualité.

⁵ Kerbrat-Orecchioni (2017, p. 22) rapproche la notion d'observable de celle de « signifiant *actualisé* (car en langue, nous dit Saussure, le signifiant est une “image mentale”, donc non observable par définition) ». Dans notre contribution, la notion d'*actualisation* voudrait donc insister sur l'appréhension de l'observable qui, par l'analyse du corpus, acquiert une matérialité.

En ce qui concerne la structure des corpus numériques, il est important de remarquer que « le corpus est avant tout sériel [...], et non nécessairement organisé linéairement » (Mayaffre 2010, p. 243). À ce propos, Mayaffre souligne que l'organisation non linéaire du corpus numérique contribue à son pouvoir heuristique. Comme « le numérique permet des lectures hypertextuelles dont la caractéristique est justement de s'affranchir du linéaire »⁶ (Mayaffre 2010, p. 244), il rend possible de nouveaux modes d'appréhension des données, favorisant également l'émergence de nouveaux observables.

La structuration des données en corpus contribue ainsi à distinguer et à matérialiser des niveaux et des paramètres d'analyse qui seront essentiels pour l'étude des observables. D'où notre proposition de considérer le corpus comme un « observatoire à observables »⁷ (Cagninelli 2023), à savoir un observatoire des actualisations potentielles de l'objet de recherche à partir de données authentiques attestées et contextualisées à plusieurs niveaux. En effet, si le corpus donne accès à un observé empirique potentiellement significatif et pertinent selon le cadre et les objectifs posés, ce sera ensuite l'analyste, par l'exploration du corpus, qui devra (re)construire les observables pertinents, en établissant les faits linguistiques et les phénomènes déterminant leur pertinence.

Ancrant notre recherche dans le cadre de l'analyse du discours française (ADF), une telle conception du corpus s'attache à faire face aux enjeux soulevés par des problématiques discursives. Ces dernières visent à étudier la matérialité linguistique à la lumière des influences exercées par les pratiques sociales et les déterminants extradiscursifs lors de leur production (cf. entre autres Maingueneau 2014). Notre approche présuppose en outre que la constitution du corpus et/ou son exploration adopte un principe d'hétérogénéité (Garric 2012 ; Garric, Longhi 2012) qui donne accès à la variation des pratiques et des usages discursifs. En l'occurrence, la structuration du corpus participe à la matérialisation de niveaux d'hétérogénéité au service d'une analyse combinant à son tour des méthodes diversifiées et complémentaires⁸. Le principe différentiel est d'ailleurs plutôt distinctif de l'exploration outillée de corpus numériques pour en identifier les traits caractéristiques : « À l'instar de toute entreprise de caractérisation, l'exploration de corpus est très souvent **contrastive**, puisqu'elle repose sur

⁶ Mayaffre (2010, p. 244) précise toutefois que « la linéarité apparaît irréductible à la textualité et constitue le socle de sa définition ».

⁷ Le développement théorique de notre proposition d'envisager le corpus comme un *observatoire à observables*, abordé en relation avec la question des observables, fera l'objet d'une étude ultérieure.

⁸ Pour une description plus ample de la conception et constitution d'un corpus hétérogène, exploré par une démarche d'analyse multidimensionnelle combinée, voir par exemple Cagninelli (sous évaluation).

des comparaisons permettant de mettre au jour des différences entre les unités choisies » (Poudat, Landragin 2017, p. 30).

Dans le cadre d'un corpus conçu pour une étude discursive, les métadonnées contribuent finalement à matérialiser des niveaux d'hétérogénéité au service d'un principe différentiel, tout comme à réfléchir sur les conditions théoriques et méthodologiques pouvant favoriser la comparabilité de données hétérogènes.

3.1. Métadonnées contextuelles et analytiques

La nature des informations intégrées aux corpus sous forme de métadonnées nous amène à proposer d'en distinguer deux classes principales ayant des répercussions différentes sur l'appréhension et l'analyse des données linguistiques, mais aussi sur les possibilités d'émergence des observables. Les métadonnées peuvent relever de deux macro-catégories d'informations : des informations empiriques de nature contextuelle aux matériaux linguistiques sélectionnés et des informations analytiques résultant de concepts théoriques ou théorico-méthodologiques.

Les *métadonnées contextuelles* se rattachent à des éléments plus objectifs et empiriques attestés au niveau de l'unité d'observation et de constitution du corpus, à savoir l'unité textuelle dans le cas d'un corpus textuel. À l'intérieur de cette catégorie, nous regroupons deux types d'informations : des paramètres situationnels-factuels qui concernent les conditions de production (par ex. temps, lieu, participants, etc.), ainsi que des paramètres formels-structurels qui déterminent l'unité textuelle dans son ensemble. De fait, les discours produits au sein d'un même contexte peuvent présenter des formats structurels différents, tout en correspondant à une même forme de textualité et à un même genre (techno)discursif, à l'instar d'un tweet-réponse qui ne partage pas les mêmes caractéristiques structurelles d'un retweet. Il s'agit donc d'informations empiriques d'ordre formel-structurel relatives au macro-niveau de l'unité d'observation, qui sont cependant susceptibles de varier en fonction tant de la forme de textualité observée que du genre du discours dont elle relève, à la différence des paramètres situationnels qui ne dépendent que du contexte de production. Si les informations situationnelles et les caractéristiques formelles-structurelles constituent des paramètres objectifs de nature contextuelle-empirique, ces deux types de métadonnées contextuelles peuvent toutefois être exploités différemment dans l'analyse, puisqu'ils peuvent servir de point de départ pour la concrétisation de différents niveaux d'analyse qui ne sont pas forcément superposables (cf. *infra*).

Nous envisageons en revanche comme *analytiques*, les métadonnées qui relèvent de catégorisations et d'annotations effectuées sur la base de notions théoriques d'ordre linguistique, discursif ou technodiscursif. Du point

de vue linguistique, l'une des annotations les plus utilisées est l'étiquetage morpho-syntaxique, qui associe à chaque forme du corpus le lemme et la catégorie grammaticale correspondants. Si ce type d'informations analytiques peut résulter aussi d'une procédure informatisée intégrée au logiciel d'analyse, cela n'empêche pas d'ajouter, manuellement ou automatiquement, d'autres informations lexico-sémantiques (à l'instar de connotations positives / négatives en association à des formes d'intérêt) et syntaxiques (concernant par exemple l'aspect du verbe ou la fonction syntaxique d'un élément dans l'énoncé) sous forme de métadonnées analytiques pour des besoins d'analyse spécifiques. La classification d'un texte comme appartenant à un genre de discours particulier constitue en revanche un exemple de métadonnées analytiques d'ordre discursif. Enfin, dans le cas des textualités numériques natives, des métadonnées analytiques technodiscursives peuvent également être prévues, à l'instar de regroupements des unités textuelles en fonction du rôle primaire ou complémentaire que la partie verbo-textuelle peut assumer dans des technodiscours tels que les tweets (voir la différence entre tweets « simples » et tweets « augmentés » abordée dans la section 4)⁹.

Les métadonnées analytiques actualisent un certain cadrage théorique sur les données sous forme de catégorisations théoriques ou d'annotations théorico-méthodologiques pouvant s'appuyer aussi sur d'autres métadonnées. La structuration du corpus par des métadonnées analytiques peut en effet se fonder sur des catégories théoriques préétablies et attestées par la littérature sur le sujet, ou bien sur des catégorisations en phase de théorisation et de vérification. L'intégration de métadonnées analytiques entraîne par conséquent l'activation d'un angle d'appréhension spécifique des données et pose les conditions pour accéder à un certain type de caractérisation et donc d'observables. Dans la mesure où ces métadonnées dépendent des critères de sélection et des notions adoptés par l'analyste, elles prennent appui sur des paramètres plus subjectifs au sens qu'ils sont susceptibles de varier en fonction de l'ancrage théorique. Par les métadonnées analytiques, les données sont ainsi inscrites dans une conception théorique qui concrétise une perspective d'appréhension et d'analyse spécifique.

Néanmoins, comme les métadonnées contextuelles permettent à leur tour de matérialiser des niveaux d'hétérogénéité au service d'une analyse contrastive-différentielle, elles peuvent aussi servir de base à l'élaboration de catégorisations théorico-méthodologiques, déterminant la création de métadonnées analytiques. Il est ainsi possible d'envisager des catégorisations théorico-méthodologiques à partir de métadonnées contextuelles d'ordre *situationnel* pour étudier des variations qui touchent aussi bien le niveau

⁹ Pour une illustration plus ample des métadonnées analytiques discursives et technodiscursives ici mentionnées, nous renvoyons à la section 4 et notamment au tableau 2.

intergénérique – à travers des formes de textualités et des genres de discours différents – que le niveau intragénérique – à travers des textes relevant d’une même forme de textualité et d’un même genre de discours. C’est le cas par exemple pour l’étude de la variation diachronique d’un fait linguistique ou bien de la variation selon différentes instances d’énonciation. Dans les deux cas, le moment de production des discours envisagés ou la source énonciative qui en est à l’origine correspondent à des paramètres objectifs¹⁰ pouvant être comparés à travers différentes formes textuelles et différents genres discursifs, bien que leur sélection soit soumise aux choix de l’analyste en fonction de l’objet et des objectifs de la recherche.

En revanche, les métadonnées contextuelles relatives au format *structurel* de la textualité peuvent servir de base pour l’élaboration de métadonnées analytiques visant l’étude de variations qui s’attestent au niveau intragénérique et non nécessairement au niveau intergénérique. Si ces métadonnées d’ordre structurel se rapportent à leur tour à des éléments empiriques propres à la matérialité de la production envisagée dans sa globalité, les informations structurelles ainsi catégorisées sous forme de métadonnées contextuelles peuvent changer en fonction aussi bien de la forme de textualité de la production discursive observée que de ses caractéristiques génériques. À titre d’exemple, on peut mentionner le cas des différents types de tweets existants qui se distinguent par des formats structurels spécifiques, tout en relevant à la fois d’une même forme de textualité et d’un même technogène de discours. Le format structurel peut ainsi constituer un paramètre factuel servant de base pour établir une catégorisation théorico-méthodologique d’ordre technodiscursif sous forme de métadonnées analytiques. En effet, si le format structurel d’un tweet peut refléter une pratique (techno)discursive spécifique, il se peut qu’il n’existe pas toujours une correspondance biunivoque entre un format structurel – à savoir un type de tweet – et la pratique technodiscursive qu’il réalise (cf. exemple 3 dans la section 4).

3.1.1. Fonctions et applications des métadonnées : vérification d’hypothèses théoriques et caractérisations linguistico-discursives

Les différentes métadonnées peuvent être exploitées pour remplir diverses fonctions au cours de l’analyse du corpus. En premier lieu, l’exploration du corpus permet de vérifier des hypothèses théoriques, en se servant aussi des catégorisations théoriques proposées sous forme de métadonnées. Comme l’affirmait Culioli déjà en 1968 : « l’ordinateur ne peut que vérifier

¹⁰ Il peut arriver qu’il ne soit pas possible de déterminer la datation exacte d’un discours ou d’en identifier la source énonciative de production. Cela n’empêche pas toutefois que ce type d’informations concerne des aspects objectifs.

l'adéquation d'une théorie en vérifiant la pertinence et la consistance d'un jeu de descripteurs, mais ne permettra jamais de faire l'économie du travail théorique » (p. 106).

La saillance de ces catégorisations attestée par l'analyse du corpus n'est pourtant que le point de départ pour caractériser ensuite les unités d'observation ainsi différenciées et comparées. À titre d'exemple, on peut mentionner l'analyse que nous avons réalisée dans un travail précédent (Cagninelli, sous évaluation) sur un corpus d'articles de journal structuré selon une distinction interne entre discours d'information et discours d'opinion, où il a été possible d'identifier des traits linguistiques sur la base desquels ressortent des divergences entre ces sous-parties. Il s'agit en l'occurrence de caractéristiques relatives à la dimension énonciative-pragmatique des productions rassemblées en corpus. Ces résultats liés à la caractérisation pragma-énonciative du corpus seraient d'ailleurs susceptibles de généralisation dans la mesure où ils relèvent de l'influence des spécificités sous-génériques sur ces productions. D'autres textes appartenant à ces mêmes sous-genres de discours pourraient ainsi présenter des caractéristiques similaires au niveau énonciatif-pragmatique. L'une des divergences montrées par cette analyse concerne en effet la présence plus ou moins significative de ressources énonciatives et modales qui participent d'une énonciation « plutôt *subjectivée* » ou « plutôt *objectivée* », pour reprendre la distinction proposée par Moirand (2007).

Afin de structurer les textes en corpus en fonction d'une hétérogénéité sous-générique et pragma-énonciative (Cagninelli, sous évaluation), nous nous sommes appuyée sur les catégorisations théoriques avancées par des études précédentes sur le sujet (Adam 2001 ; Moirand 2007 ; Charaudeau 2011), que nous avons prises comme référence. L'analyse effectuée sur le corpus ainsi structuré nous a permis de tester les critères et les hypothèses de classification des textes en fonction de ces deux sous-genres du discours journalistique, en clarifiant les ressources linguistiques qui en sont distinctives. Les résultats confirment et enrichissent les catégorisations théoriques existantes et ils seront utiles lors de comparaisons ultérieures pour tester l'évolution de ces traits distinctifs.

3.1.2. Fonctions et applications des métadonnées : création des conditions d'observation de nouveaux observables

Le corpus enrichi de métadonnées assure différentes perspectives d'observation, contribuant à faire ressortir des éléments distinctifs et des phénomènes qui n'avaient même pas fait l'objet d'hypothèses au stade de sa conceptualisation. L'exploration outillée du corpus exploitant les niveaux d'observation et de comparaison matérialisés par les métadonnées permet en effet d'accéder à de nouveaux observables, qui seraient difficilement

perceptibles autrement (cf. aussi Rastier 2011). Il nous semble important d'insister en l'occurrence sur l'apport heuristique du recours à des outils d'exploration informatisée (Mayaffre 2002, 2005) qui favorisent des approches variées des données (Garric 2012). Ainsi, « l'ordinateur se révèle [...] un outil heuristique susceptible de reculer l'horizon de nos investigations » (Mayaffre 2002, p. 159).

En effet, ce sont les diverses formes et modalités d'appréhension et d'observation des données par le biais de l'outil informatique qui transforment le rapport de l'analyste aux données. Précisons-le tout de suite : il ne s'agit nullement de dévaloriser l'approche qualitative des données, bien au contraire. L'analyse qualitative et la démarche interprétative de l'analyste se nourrissent des observations rendues possibles par le recours à l'outil, de même que la réflexion théorique informe le corpus en concrétisant un cadrage spécifique de la matérialité langagière. L'analyste joue donc un rôle essentiel non seulement au cours de l'exploration du corpus, mais aussi lors de sa conception et de sa constitution (cf. aussi Pincemin 2020). Une exploration informatisée qui s'effectue sans sélectionner les outils et les fonctions adéquates aux données et aux objectifs envisagés, de même qu'une exploration qui ne se modèle pas au fur et à mesure que l'étude avance et que de nouveaux questionnements surgissent, peut au contraire se révéler stérile. Les apports de l'exploration outillée se rattachent au rôle de l'analyste tant dans les méthodes adoptées que dans l'interprétation et la mise en relation des résultats obtenus afin d'accéder à des observables pertinents dont l'analyse permet aussi des retombées théoriques.

3.2. Métadonnées de décomplexification

La structuration et l'analyse d'un corpus rassemblant des textualités natives de l'environnement numérique posent des défis supplémentaires, notamment si l'on envisage ces productions comme des « technodiscours » (Paveau 2017), sans pour autant se priver des apports d'une approche outillée. Afin de saisir les technodiscours dans leur complexité, la matérialité langagière dont ils se composent gagne à être appréhendée en relation avec leurs constituants non-langagiers et avec l'environnement natif de production, selon une perspective « écologique » (Paveau 2017).

D'un point de vue méthodologique, il s'avère donc nécessaire de saisir la nature complexe et composite des technodiscours (voir la section 2) à travers un processus de décomplexification initiale qui permet de distinguer les divers constituants dont dépendent au moins en partie leurs propriétés spécifiques, avant de les mettre en relation entre eux. Par conséquent, un corpus constitué de technodiscours requiert une classe supplémentaire de métadonnées pour individualiser ses composantes et les associer à la matérialité linguistique de ces productions : les métadonnées de

décomplexification. À la différence des métadonnées contextuelles qui concernent le niveau macro-textuel de l'unité textuelle retenue comme unité de base du corpus, les métadonnées de décomplexification rendent compte d'éléments constitutifs de cette unité, en décomposant la complexité de la production envisagée dans sa globalité pluritechnosémiotique. La mise au point et l'adoption de métadonnées de décomplexification relève ainsi d'une approche spécifique des textualités complexes de nature pluritechnosémiotique. La dimension linguistique de ces productions reste en l'occurrence au centre de l'attention de l'étude tout en étant observée eu égard aux spécificités technodiscursives distinctives. À titre d'exemple, la dimension relationnelle et l'augmentation énonciative impliquées par les ressources technolangagières (cf. Paveau 2017), telles que les mentions utilisées dans les posts des RSN, constituent l'une des composantes technodiscursives dont la prise en compte permet d'aborder la matérialité langagière et, plus généralement, le sens de ces productions de manière plus complète voire plus complexe (pour une plus ample illustration, voir la section 4).

En restituant les différentes composantes des technodiscours, ces métadonnées contribuent à créer les conditions pour une observation et pour des analyses outillées plus fines et complètes, qui doivent toutefois se combiner avec des démarches contextualisées et avec le retour au technodiscours dans son contexte de production (cf. aussi Longhi 2020, 2021 ; Cagninelli 2022). De fait, l'individualisation des éléments constitutifs des technodiscours intégrés sous forme de métadonnées représente une étape nécessaire pour outiller l'analyse et pour accéder enfin à la complexité de l'observé d'origine. La déstructuration de celui-ci répond ainsi à une nécessité épistémologique et méthodologique qui sert d'appui pour atteindre une approche plus exhaustive.

Dans ce cas aussi, les objectifs de la recherche déterminent les métadonnées de décomplexification à intégrer au corpus. Pour ce faire, une bonne connaissance des matériaux sélectionnés est néanmoins indispensable. Il faut connaître non seulement les spécificités technodiscursives des textualités envisagées, mais aussi leurs propriétés formelles et technologiques, qui sont également liées au format des données retenues. À cet égard, on peut mentionner la distinction posée par Paveau (2013) entre les deux perspectives méthodologiques principalement adoptées pour aborder les discours natifs du web 2.0, qui privilégient respectivement l'« extraction » ou la « contextualisation » du matériau linguistique. Chacune s'appuie, du fait de l'approche adoptée, sur des formats et des traitements informatiques différents des données linguistiques ainsi que d'autres éléments constitutifs des technodiscours pouvant être éventuellement pris en compte. Plus récemment, des études ont par ailleurs proposé d'articuler ces deux approches afin d'aborder ces textualités de manière plus complexe, comme dans le cas

des travaux de Longhi (2020, 2021). L'élaboration de métadonnées de décomplexification et leur application lors de l'exploration du corpus s'inscrivent ainsi dans le sillage de ces études visant une approche outillée plus complexe des discours numériques, où les analyses d'ordre statistique et mesurable sont effectuées et interprétées en relation avec l'observation en contexte du matériau linguistique traité informatiquement (cf. aussi Cagninelli 2022).

À l'instar des métadonnées contextuelles, les métadonnées de décomplexification peuvent également servir de base pour l'élaboration d'annotations et de catégorisations nouvelles des données. En d'autres termes, à partir des métadonnées de décomplexification, l'analyste peut établir des métadonnées analytiques, contribuant à encadrer théoriquement les données et/ou à introduire de nouveaux niveaux d'analyse et de caractérisation. L'ajout de ces annotations peut relever donc d'« une activité interprétative du chercheur. L'annotation s'apparente ainsi à ce que le fait le linguiste tous les jours : analyser, décrire, classer le matériau langagier » (Née, Fleury 2017, p. 85). Il faut toutefois préciser que les annotations auxquelles nous nous référons ici ne correspondent ni aux annotations morphosyntaxiques ou sémantiques (Née, Fleury 2017), ni aux annotations complexes (Poudat, Landragin 2017). Elles matérialisent plutôt des classements d'éléments au niveau des unités textuelles ou des unités d'observation sur la base de l'un ou plusieurs des paramètres ainsi concrétisés. Les métadonnées de décomplexification et les annotations qui s'ensuivent (pour illustration voir *infra* Tableau 2) donnent corps à plusieurs formes de mise en contexte des matériaux verbaux, enrichissant ainsi l'interprétation.

4. Corpus numériques et métadonnées : exemples de structuration et d'application pour l'analyse

Nous proposons ici quelques exemples de métadonnées que l'on pourrait envisager dans le cadre d'un corpus de tweets¹¹, représentant une forme de textualité que nous avons déjà étudiée¹² (parmi d'autres, voir Cagninelli 2022, 2024). En ce qui concerne les métadonnées contextuelles, on pourrait

¹¹ Rappelons que, à la suite du changement de nom de la plateforme Twitter en X en juillet 2023, les tweets ont été renommés « posts ». Comme nous l'avons précisé plus haut, nous utiliserons en l'occurrence les dénominations employées à l'époque où remontent les exemples sur lesquels prend appui notre illustration de quelques cas d'analyse.

¹² Comme nous l'avons déjà souligné, l'élaboration des métadonnées analytiques et de décomplexification requiert une bonne connaissance de la textualité – ou des textualités – dont relèvent les données du corpus. D'où la nécessité d'illustrer notre propos à partir de textualités qui ont déjà fait l'objet de nos recherches.

par exemple retenir des paramètres temporels à l’instar de la date de publication. Cela permettrait d’observer la variation diachronique en tenant compte également d’éventuels événements marquant la période sélectionnée. Comme dans tout partitionnement, l’unité sélectionnée pour segmenter le corpus pourra toutefois avoir des répercussions sur le niveau de granularité des résultats (cf. aussi Rastier 2011). Une autre catégorie de métadonnées contextuelles pourrait correspondre au type de dispositif (un ordinateur ou un portable) employé pour la création du message afin d’analyser l’influence du support matériel sur les caractéristiques linguistiques de la production technodiscursive.

Relativement aux métadonnées analytiques, on pourrait en revanche supposer des catégorisations qui relèvent de notions théoriques transversales aux formes de textualités, à l’instar d’un partitionnement du corpus en sous-parties d’ordre générique et sous-générique. En l’occurrence, la création de métadonnées analytiques pourra procéder aussi d’annotations ou de catégorisations élaborées à partir de métadonnées contextuelles (comme dans le cas de la catégorisation des différentes pratiques technodiscursives à partir des types de tweets distingués en fonction de leur format structurel ainsi que de leur dimension pragma-énonciative, cf. *infra* Tableau 2 et exemple 3) et notamment de métadonnées de décomplexification. Celles-ci peuvent rendre compte d’éléments constitutifs des tweets à l’instar de la présence de ressources technolangagières spécifiques telles que les hashtags et les mentions, la présence d’un lien, d’une image ou d’autres contenus médias, ou encore des taux d’engagement relatifs au nombre de réponses, de retweets, de citations et ainsi de suite, pour ne citer qu’eux. Toutes ces informations constitutives pourraient donc devenir de nouveaux paramètres à prendre en considération – seuls ou en combinaison entre eux – pour caractériser la matérialité langagière des technodiscours et examiner les variations qui s’ensuivent.

Ce sera néanmoins la problématique de la recherche qui déterminera le type de métadonnées de décomplexification à envisager ou encore l’ajout d’annotations supplémentaires à partir de celles-ci. L’analyste devra définir et délimiter ces catégorisations qui s’appuient sur des éléments constitutifs attestés. Les métadonnées de décomplexification jouent ainsi un rôle de premier plan dans la déconstruction initiale de la complexité technodiscursive et dans sa reconstruction au fil de l’analyse.

Afin de donner une illustration concrète de la façon dont des métadonnées de décomplexification peuvent servir de base pour l’élaboration de métadonnées analytiques apportant une catégorisation théorique – déjà attestée dans la littérature ou en cours d’élaboration – sur les données empiriques, nous revenons sur la distinction que nous avons proposée entre tweets « simples » et tweets « augmentés » (Cagninelli 2022). Cette distinction vise à rendre compte de deux macro-catégories de structuration

pluritechnosémiotique du tweet en fonction de la présence ou absence d'éléments qui augmentent sa sémiologie, à l'instar des liens URL et des différents types de contenus visuels ou audiovisuels pouvant être intégrés dans les tweets. Elle prend appui sur une réflexion théorique concernant le rôle du composant verbo-textuel – c'est-à-dire le message linguistique affiché dans la partie supérieure du tweet – dans la création du sens de l'ensemble du technodiscours.

Notre hypothèse était que les tweets qui ne comportent qu'une partie verbo-textuelle – que nous avons appelés tweets « simples » – pourraient généralement être plus autosuffisants sur le plan sémantique et sémiotique, puisque leur composante langagière n'entre pas en relation avec d'autres éléments sémiotiques pour la création du sens global. Au contraire, la composition plurisémiotique des tweets augmentés pourrait plus fréquemment aboutir à des contenus verbo-textuels sémantiquement incomplets, mais dont le sens est saturé par d'autres contenus sémiotiques. Cette hypothèse de catégorisation théorique a été matérialisée sous forme de métadonnées analytiques – par l'opposition tweets simples / tweets augmentés – à partir de métadonnées de décomplexification rendant compte de la présence de contenus sémiotiques supplémentaires. Les tweets dont les métadonnées de décomplexification attestent la présence de liens, de contenus iconiques et audiovisuels ont ainsi servi comme point de départ pour la création de la macro-catégorie des tweets « augmentés ». Par ricochet, les métadonnées de décomplexification signalant l'absence de contenus sémiotiques supplémentaires au texte ont permis de classer les tweets « simples ». L'exploration du corpus effectuée sur la base de cette distinction a enfin fait ressortir des divergences entre les deux macro-catégories, permettant de vérifier l'hypothèse en corpus.

Le tableau 2 offre une synthèse des exemples des métadonnées que l'on vient d'illustrer, en les classant en fonction de la typologie avancée. Il montre en outre les croisements entre les métadonnées contextuelles ou de décomplexification, d'une part, et les métadonnées analytiques, de l'autre.

Classes de métadonnées	Types d'information	Valeurs ou catégorisations correspondantes
Métadonnées contextuelles	Informations sur le contexte de production et sur le format structurel du tweet	Date
		Dispositif
		Type de tweet
Métadonnées de décomplexification (constituants du technodiscours)	Constituants pluritechnosémiotiques	Présence/absence de liens
		Présence/absence d'images
		Présence/absence de vidéos
		Présence/absence de GIFs, etc.
	Constituants technolangagiers	Présence/absence de hashtags
		Présence/absence de mentions
	Taux d'engagement	Nombre de retweets
		Nombre de citations
		Nombre de <i>j'aime</i>
		Nombre de réponses
Métadonnées analytiques (catégorisations théoriques)	Genre discursif	Technodiscours – tweet
	Structuration pluritechnosémiotique (annotation des métadonnées de décomplexification)	Tweets simples vs tweets augmentés
	Spécificités pragma-énonciatives (annotation des métadonnées contextuelles)	Pratiques technodiscursives initiatives, réactives, enchainâtes, citationnelles, etc.

Tableau 2

Exemplification de quelques métadonnées envisageables pour un corpus de tweets selon la typologie proposée.

Afin de mettre en pratique cette typologie, nous montrons des applications possibles à partir d'une sélection d'exemples. Pour chacun, nous rendons compte de plusieurs métadonnées envisageables et classifiables selon les trois classes proposées ci-dessus. Compte tenu de la nature illustrative du propos, les métadonnées ne seront donc pas sélectionnées en vue d'un objectif d'analyse déterminé. Il convient néanmoins de souligner de nouveau que la nature des données à examiner et l'objet d'étude envisagé représentent des éléments fondamentaux pour la constitution, la structuration et l'analyse de tout corpus.

L'exemple 1 illustre un tweet du compte Twitter – renommé X depuis juillet 2023 – d'Emmanuel Macron, dont le nom d'utilisateur correspondant – appelé aussi *pseudonyme* ou *pseudo* – est @EmmanuelMacron, publié le 12 décembre 2020.



Figure 1
Tweet d'Emmanuel Macron.

Au cas où ils seraient pertinents pour les objectifs de l'étude et pour la mise en place d'analyses contrastives, ces trois types d'informations – le nom du compte, le pseudonyme et la date – peuvent être retenus et catégorisés sous forme de métadonnées contextuelles. En ce qui concerne le format structurel du tweet, il est ensuite possible de classer l'exemple 1 comme un tweet « original ». Sur la base de son format et de sa nature énonciative et pragmatique, ce tweet peut ensuite être catégorisé du point de vue théorico-méthodologique comme une pratique technodiscursive à valeur initiative, dans la mesure où il lance la conversation – à la différence des tweets-réponses et des retweets avec ou sans citation qui procèdent d'un tweet antérieur – sans qu'il soit en même temps prolongé par des tweets consécutifs et enchaînés les uns aux autres, comme c'est le cas pour les threads.

Quant aux constituants des technodiscours qui peuvent donner lieu à des métadonnées de décomplexification, on peut renvoyer, d'une part, à la présence des « technomots » (Paveau 2017) – en l'occurrence, un hashtag (*#5AnsAprès*) – et, de l'autre, aux taux d'engagement du tweet. Les taux d'engagement sont affichés dans la partie inférieure de chaque tweet par le biais d'icônes associées à des valeurs chiffrées. La petite bulle renvoie en effet au nombre de réponses obtenues par le tweet (327), les deux flèches courbes pointant respectivement vers le haut et vers le bas indiquent le nombre de fois que le tweet a été rediffusé sur la plateforme via le retweet (415), alors que l'icône du cœur correspond au nombre de réactions obtenues via le bouton du « j'aime » (2 k). Étant donné que le tweet n'inclut pas d'autres contenus sémiotiques augmentant la partie verbo-textuelle, il est possible de le catégoriser comme un *tweet simple*, selon la distinction mentionnée plus haut, en relation avec les métadonnées analytiques concernant la structuration pluritechnosémiotique.

L'exemple 2 consiste en un tweet du compte du ministère de l'Économie et des Finances – @Economie_Gouv – du 8 décembre 2020. Au-delà des éléments qui peuvent faire l'objet de métadonnées et donc de catégorisations similaires ou superposables au cas précédent (nom du compte, pseudo, date et format du type tweet « original » pour les métadonnées contextuelles ; pratique technodiscursive initiative et structuration du type « tweet simple » pour les métadonnées analytiques ; technomots et taux d'engagement pour les

métadonnées de décomplexification), l'attention se dirige en l'occurrence sur les constituants qui présentent en revanche des différences.



Figure 2
Tweet du ministère de l'Économie et des Finances.

En ce qui concerne les constituants technolangagiers, on remarque la présence de plusieurs hashtags (*#JournéeMondialeDuClimat*, *#climatique*, *#environnement*, *#PlanClimat*) ainsi que d'une mention (*@Gouvernement*). Les hashtags et les mentions constituent deux types de métadonnées de décomplexification distincts, pouvant ou non être regroupés dans une seule classe de métadonnées analytiques en fonction de leurs propriétés technolangagières, selon les exigences de l'étude. Deux autres éléments constitutifs du tweet et ici présents¹³ peuvent correspondre à autant de métadonnées de décomplexification, à savoir l'inclusion d'un lien URL et celle d'un contenu iconique. En suivant le tableau précédent, ces deux métadonnées de décomplexification peuvent servir de base pour la création de la catégorie des *tweets augmentés*, relevant de métadonnées analytiques.

Le dernier exemple nous permet enfin d'attirer l'attention sur la catégorisation du type de tweet et de la pratique technodiscursive qu'il réalise, les deux pouvant ne pas correspondre de manière biunivoque.

¹³ Il est possible de prévoir ces mêmes métadonnées pour l'exemple 1 aussi. Il est néanmoins évident que dans ce cas les deux catégories signaleraient l'absence de liens et de contenus iconiques.



Figure 3
Tweet-réponse d'Emmanuel Macron.

Du point de vue formel-structurel et donc du type de tweet associé, l'exemple 3 constitue un tweet-réponse, comme l'indique l'expression *En réponse à* affichée juste au-dessous des informations de contextualisation. On remarque toutefois que le pseudonyme auquel la réponse est adressée coïncide avec celui du compte qui est à son origine. Il s'agit ainsi d'un cas particulier qui, tout en correspondant formellement à une réponse, mérite une réflexion lors de sa catégorisation par les métadonnées analytiques. Il est en effet utile de rappeler que cette pratique de répondre à soi-même était utilisée pour développer son propos à travers plusieurs tweets, avant que la plateforme n'introduise une fonctionnalité spécifique : le thread. Comme le remarque Bibié (2019), cet usage détourné de la réponse permettait de contourner la contrainte des 180 caractères par message prévue par la plateforme¹⁴. Dans la mesure où cette forme d'autoréponse s'apparente pragmatiquement à la pratique technodiscursive incarnée par les threads, il reviendra à l'analyste d'évaluer s'il est plus pertinent pour son étude de classer cette occurrence comme une réponse ou comme un thread.

On remarque ainsi qu'à la fois les métadonnées contextuelles et les métadonnées de décomplexification peuvent servir de base pour l'élaboration de nouvelles catégorisations à annoter sous forme de métadonnées analytiques. En l'occurrence, la métadonnée analytique de nature technodiscursive identifie l'une de deux pratiques technodiscursives (réactive ou enchaînante) dont relève le tweet, en prenant appui, d'une part, sur une métadonnée contextuelle qui catégorise le type structurel du tweet envisagé et, de l'autre, sur l'interprétation de sa nature pragmatique de la part de l'analyste en fonction des exigences et des objectifs de l'étude.

Le tableau 3 met en relation la typologie de métadonnées proposée et l'application qui en a été illustrée à partir des exemples 1-3.

¹⁴ Des changements concernant la limite des caractères par tweet – augmentée à 180 en 2017 – ont été également apportés depuis 2023.

Classes de métadonnées Type d'information et catégorisations associées		Valeurs associées aux catégories de métadonnées à partir des exemples 1-3		
Métadonnées contextuelles		<i>Exemple 1</i>	<i>Exemple 2</i>	<i>Exemple 3</i>
Contexte de production	Nom du compte	Emmanuel Macron	Ministère de l'Économie et des Finances	Emmanuel Macron
	Pseudo	@EmmanuelMacron	@Economie Gouv	@EmmanuelMacron
	Date	12/12/2020	08/12/2020	12/12/2020
Format structurel	Type de tweet	Tweet « original »	Tweet « original »	Tweet-réponse
Métadonnées de décomplexification		<i>Exemple 1</i>	<i>Exemple 2</i>	<i>Exemple 3</i>
Constituants pluritechno-sémiotiques	Liens URL	Absents	Lien URL présent	Absents
	Images	Absents	Contenu icono-textuel présent	Contenu icono-textuel présent
Constituants technolangagiers	Hashtags	#5AnsAprès	#JournéeMondiale DuClimat #climatique #environnement #PlanClimat	Absents
	Mentions	Absents	@Gouvernement	@EmmanuelMacron ¹⁵
Taux d'engagement	Retweets	415	23	50
	J'aime	2k	28	267
	Réponses	327	7	19
Métadonnées Analytiques		<i>Exemple 1</i>	<i>Exemple 2</i>	<i>Exemple 3</i>
Classification générique		Technodiscours – tweet	Technodiscours – tweet	Technodiscours – tweet
Structuration pluritechno-sémiotique	Tweet simple / augmenté	Tweet simple	Tweet augmenté	Tweet augmenté
Spécificités pragma-énonciatives	Pratique techno-discursive	Pratique initiative du type « tweet original »	Pratique initiative du type « tweet original »	Pratique réactive du type « tweet-réponse »
				Pratique enchaînante assimilable au « thread »

Tableau 3

Tableau résumptif de la catégorisation des métadonnées appliquées aux exemples.

¹⁵ À la différence de la mention présente dans l'exemple 2 qui est insérée en phase d'écriture du message, la mention de l'exemple 3 est générée automatiquement par l'acte de répondre à un tweet antérieur en cliquant le bouton respectif.

Pour résumer, dans le cadre d'un corpus de technodiscours, les différentes classes de métadonnées contribuent à traduire en corpus la complexité des matériaux à analyser et à enrichir leur appréhension. Si les métadonnées décomposent la complexité pour permettre de la saisir, ce sera ensuite à l'analyse et à l'analyste de la reconstituer par la synthèse des résultats obtenus et des interprétations avancées. D'où l'importance de la démarche d'analyse adoptée. Comme le met en évidence Longhi (2020) à partir de la distinction identifiée par Paveau (2013) et citée plus haut (cf. section 3.2), l'analyse des corpus de discours numériques natifs gagne à rechercher un équilibre entre des démarches centrées plutôt sur l'« extraction », qui se focalisent sur la matérialité linguistique, et des démarches plutôt « contextualisantes », qui visent l'appréhension du langagier en relation avec le technologique et l'environnement de production. C'est dans cette direction que vont la notion de « corpus réfléchi » de Longhi (2021) et le modèle théorique du logiciel *Visaneco* (Cagninelli, Taglioli 2022), présenté dans Cagninelli (2022). Une analyse plus fine et plus riche des technodiscours dépend donc aussi bien des conditions d'exploration créées par le corpus et par sa structure que des approches et des méthodes d'analyse adoptées.

5. Conclusion

Considérant que la réflexion théorico-méthodologique préside à la constitution du corpus et qu'elle en est enrichie en retour, la structuration du corpus représente une étape significative de la recherche, d'abord réflexive et ensuite appliquée. Les métadonnées s'avèrent en effet essentielles non seulement pour contextualiser les données, mais surtout pour introduire des niveaux d'analyse qui diversifient les angles et les méthodes pour les appréhender. La phase de structuration du corpus se pose ainsi à l'articulation entre la nature des données, le cadrage théorico-méthodologique de l'analyste et l'objet d'étude.

La classification des métadonnées avancée ici vise à attirer l'attention aussi bien sur la nature attestée ou théorique des paramètres adoptés que sur les spécificités des matériaux observés. S'appuyant sur les métadonnées en tant que paramètres d'analyse, l'exploration outillée du corpus permet d'accéder à de nouvelles configurations des données. Celles-ci offrent la possibilité de varier et de multiplier les perspectives d'approche des données, contribuant à soulever de nouvelles hypothèses. Dans le cadre d'une étude discursive, les métadonnées participent en outre du travail interprétatif mené par l'analyste dans la mesure où elles favorisent l'ancrage matériel des données linguistiques empiriques. L'exploration du corpus devra par ailleurs être également étayée par les compétences et les connaissances encyclopédiques de l'analyste, afin d'articuler les différentes perspectives

d'appréhension des données en corpus en les envisageant selon une approche discursive qui met en relation le linguistique et l'extralinguistique.

La structuration du corpus par le biais des métadonnées favorise finalement une appréhension plus fine et plus riche des matériaux à observer et alimente la réflexion et les questionnements non seulement sur l'objet d'étude mais aussi sur le cadrage théorique et sur la démarche d'analyse. En association avec l'outillage de l'analyse, les métadonnées contribuent à la portée heuristique de l'exploration outillée de corpus numériques.

Bionote : Claudia Cagninelli is post-doctoral researcher at the Department of Languages, Literatures, Cultures and Mediations of the University of Milan La Statale. She holds a PhD in Human Sciences/*Scienze del linguaggio* under the joint supervision of the University of Modena and Reggio Emilia and CY Cergy Paris University. Her research interests lie in the field of digital discourse analysis and institutional discourse analysis, with a particular focus on the semantic, enunciative and argumentative dimensions. Additionally, she is interested in theoretical and methodological issues related to the role of the corpus and computerized tools for discourse analysis.

Adresse de l'auteure : claudia.cagninelli@unimi.it

Références bibliographiques

- Adam J.-M. 1997, *Unités rédactionnelles et genres discursifs : cadre général pour une approche de la presse écrite*, in « Pratiques » 94 [1], pp. 3-18.
- Bibié L. 2019, *Utilisation de ok sur Twitter, entre (re)présentation de soi et stabilisation énonciative*, in « Lexique » 25, pp. 57-75.
- Cagninelli C. 2022, *Positionnements discursifs dans le débat public sur la fin de vie : variations génériques entre (inter)subjectivité et interdiscursivité*, Thèse de doctorat, Università di Modena e Reggio Emilia/CY Cergy Paris Université, Modena.
- Cagninelli C. 2023, *Corpus et observables discursifs : quelle(s) articulation(s) entre réflexion épistémologique et exigences méthodologiques ? Quelques considérations à partir d'un cas concret d'application*, communication présentée au Colloque jeunes chercheuses et chercheurs de R2DIP 2023, *Observables, observant.es, observés : construire et analyser des objets en analyse du discours*, 18 et 19 décembre, CY Cergy Paris Université.
- Cagninelli C. 2024, *Effets rhétoriques de l'augmentation énonciative des discours numériques. Le cas des "retweets avec citation"*, in « Lingue e Linguaggi » 62, pp. 205-226.
- Cagninelli C. sous évaluation, *Genres de discours et hétérogénéités en corpus. Une approche méthodologique des variations intra- et intergénériques*.
- Cagninelli C. et Taglioli M. 2022, *Visaneco* [Computer software].
- Charaudeau P. 2011, *Les médias et l'information. L'impossible transparence du discours*, De Boeck Université, Bruxelles.
- Condamines A. 2005, *Linguistique de corpus et terminologie*, in « Langages » 157 [1], pp. 36-47.
- Condamines A. 2022, *Présentation*, in « Éla. Études de linguistique appliquée » 208 [4], pp. 391-394.
- Constantin de Chanay H. et Ferron S. 2017, *Introduction*, in « Le discours et la langue. Revue de linguistique française et d'analyse du discours » 9 [2], pp. 7-19.
- Cori M., David S. et Léon J. 2008, *Présentation : éléments de réflexion sur la place des corpus en linguistique*, in « Langages » 171 [3], pp. 5-11.
- Culioli A. 1968, *La formalisation en linguistique*, in « Cahiers pour l'analyse » 9, pp. 106-109.
- Dalbera J.-Ph. 2002, *Le corpus entre données, analyse et théorie*, in « Corpus » 1. <http://journals.openedition.org/corpus/10>
- Duteil-Mougel C. 2006, *Groupements de textes et corpus : point de vue linguiste*, in Duteil-Mougel C. et Foulquié B. (éds.), *Actes du Colloque international et école d'été « Corpus en Lettres et Sciences sociales – Des documents numériques à l'interprétation »*, pp. 205-216.
- Garric N. et Longhi J. 2012, *L'analyse de corpus face à l'hétérogénéité des données : d'une difficulté méthodologique à une nécessité épistémologique*, in « Langages » 187 [3], pp. 3-11.
- Garric N. 2012, *Construire et maîtriser l'hétérogénéité par la variation des données, des corpus et des méthodes*, in « Langages » 187 [3], pp. 73-92.
- Kerbrat-Orecchioni C. 2017, « *Observable* » et « *observer* » en sciences du langage, in « Le discours et la langue. Revue de linguistique française et d'analyse du discours » 9 [2], pp. 21-45.
- Longhi J. 2018, *Du discours comme champ au corpus comme terrain : contribution méthodologique à l'analyse sémantique du discours*, L'Harmattan, Paris.

- Longhi J. 2020, *Explorer des corpus de tweets : du traitement informatique à l'analyse discursive complexe*, in « Corpus » 20. <http://journals.openedition.org/corpus/4567>
- Longhi J. 2021, *Du corpus réflexif au corpus réfléchi : la plateforme #Idéo2017 pour extraire contextuellement les pratiques citationnelles et analyser la circulation des discours politiques sur Twitter*, in « Le discours et la langue. Revue de linguistique française et d'analyse du discours » 12 [2], pp. 99-113.
- Longhi J. et Sarfati G.-E. 2018, *Conception du corpus et méthodologie d'analyse : Pour un renouvellement de l'analyse des discours institutionnels et politiques*, in « Semiotica » 223, pp. 87-110.
- Maingueneau D. 2014, *Discours et analyse du discours : une introduction*, A. Colin, Paris.
- Mayaffre D. 2002, *L'Herméneutique numérique*, in « L'Astrolabe. Recherche littéraire et Informatique », numéro spécial, pp. 151-161. <hal-00586512>.
- Mayaffre D. 2005, *Rôle et place des corpus en linguistique : réflexions introductives*, in « Actes des Journées d'Etude TOULOUSAINES JETOU 2005 », pp. 5-17
- Mayaffre D. 2010, *Corpus et web-corpus. Réflexion sur la corporalité numérique*, in « Cahiers de praxématique » 54-55, pp. 233-248.
- Moirand S. 2007, *Les discours de la presse quotidienne : observer, analyser, comprendre*, Presses universitaires de France, Paris.
- Née É. et Fleury S. 2017, *Constituer un corpus en trois scénarios*, in Née É. (éd.), *Méthodes et outils informatiques pour l'analyse des discours*, Presses universitaires de Rennes, Rennes, pp. 63-101.
- Neveu F. 2016, *Observatoires et observables en linguistique française*, in « Le français moderne – Revue de linguistique française » 84 [1], pp. 1-12
- Paveau M.-A. 2013, *Analyse discursive des réseaux sociaux numériques* [Dictionnaire]. *Technologies discursives*. Consulté à l'adresse <https://doi.org/10.58079/uoww>
- Paveau M.-A. 2015, *Ce qui s'écrit dans les univers numériques. Matières technolangagières et formes technodiscursives*, in « Itinéraires » 2014-1. <http://journals.openedition.org/itineraires/2313>
- Paveau M.-A. 2017, *L'analyse du discours numérique : dictionnaire des formes et des pratiques*, Hermann, Paris.
- Pincemin B. 1999, *Construire et utiliser un corpus : le point de vue d'une sémantique textuelle interprétative*, in Condaminas A., Péry-Woodley M.-P. et Fabre C. (éds), *Atelier Corpus et TAL : pour une réflexion méthodologique* (TALN 99), Cargèse, pp. 26-36.
- Pincemin B. 2012, *Hétérogénéité des corpus et textométrie*, in « Langages » 187 [3], pp. 13-26.
- Pincemin B. 2020, *La textométrie en question*, in « Le français moderne – Revue de linguistique française » 88 [1], pp. 26-43.
- Poudat C. et Landragin F. 2017, *Explorer des données textuelles : méthodes – pratiques – outils*, De Boeck supérieur, Louvain-la-Neuve.
- Rastier F. 2001, *Sémiotique et sciences de la culture*, in « Linx » [En ligne] 44. <http://journals.openedition.org/linx/1058>
- Rastier F. 2011, *La mesure et le grain. Sémantique de corpus*, H. Champion, Paris.

WEBBASIERTE TOOLS FÜR DIE TRANSKRIPTION UND ANALYSE VON REDEN

Hilfreiche Instrumentarien für die (Polito)Linguistik

MARCELLA PALLADINO
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Abstract – To conduct a linguistic analysis of political speeches, it is essential to obtain their transcription. However, not all speeches delivered by political representatives are readily transcribed and accessible. Moreover, transcribing speeches that last more than some minutes is time-consuming and for this reason written texts, short political speeches, or speeches that already have an official transcription are more likely to be considered by linguistic disciplines. Tools and technology developed by the speech processing community facilitate the transcription process. The audio signal tends to be neglected in some linguistic disciplines, which focus exclusively on the content or the argumentative structures of speeches, without considering the relevant role of voice characteristics in the success or effectiveness of a speaker. In fact, this often happens in politolinguistics and this paper presents a possible alternative path for transcription of political speeches with BAS web tools, based on previous studies carried out for transcription in other disciplines. The use of such tools opens the perspective of including the signal in the analyses. The corpus of speeches used as examples was collected for a doctoral project on political speeches in Germany and Italy. The aim of this study is not to outline a technical perspective on the development and commands of transcription tools or software, but to show how to solve common problems that researchers in political language could face, once they want to consider for instance longer speeches delivered at political conventions and not transcribed before. Based on the problems and solutions that have been found for the doctoral project, researchers in the (polito)linguistic field could also determine whether the presented tools are applicable for their research and perhaps useful to make time-consuming tasks as transcribing easier. The article will present such tools, discuss their advantages and disadvantages, and conclude with examples and recommendation of best practices.

Keywords: political language; transcription; web tools; politolinguistics; audio signal.

1. Einleitung

Dieser Beitrag befasst sich mit den sogenannten BAS Web Tools¹, die bei der Transkription bzw. Korrektur von Transkripten² und der Analyse von Reden eine wertvolle Hilfe darstellen können. Anhand von Beispielen aus einem

¹ <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface> (25.04.2023).

² Der Terminus ‚Transkription‘ bezeichnet den Prozess, wohingegen das ‚Transkript‘ das Ergebnis der Transkription ist.

deutsch-italienischen Korpus politischer Reden werden Wege vorgestellt, die den Prozess des Transkribierens und die Analyse des Audiosignals erleichtern. Der Beitrag basiert auf den konkreten Erfahrungen mit der Transkription für ein Korpus politischer Reden im Rahmen eines Doktorats und hat nicht die Absicht technische Anleitungen zu liefern. Das Ziel ist vielmehr, einige Tools vorzustellen, die für ForscherInnen aus der (Polito)Linguistik interessant sein können (für eine ausführliche Definition der Disziplin siehe u.a. Burkhardt 1996; Niehr 2014). In einer Pilotstudie konnte Christoph Draxler (PD am Institut für Phonetik und Sprachverarbeitung an der Ludwig-Maximilians-Universität in München) kritische Aspekte des Transkriptionsprozesses, zum Beispiel die Wortfehlerrate und die Transkriptionsgeschwindigkeit, auf der Grundlage der hier untersuchten Reden herausfiltern und analysieren (Draxler 2023). Allerdings sind die Ergebnisse noch vorläufig und eine weitere Erforschung des Themas ist wünschenswert.

Aufgebaut ist der Artikel wie folgt: Nach einer theoretischen Einführung zum Thema Transkription in sprachwissenschaftlichen Disziplinen wie der Politolinguistik (Abschnitt 2) werden ausgewählte BAS Web Tools vorgestellt (Abschnitt 3). Anhand des für das Promotionsprojekt erstellten Korpus wird eine praktische Anwendung der BAS Tools gezeigt (Abschnitt 4), bevor abschließend (Abschnitt 5) ein Fazit gezogen wird.

Das Thema des Beitrags knüpft folglich an die Sektion dieser Sonderausgabe an, die sich mit den digitalen Transformationen in diversen linguistischen Bereichen und den daraus resultierenden Implikationen beschäftigt. Webbasierte BAS Tools stellen eine neuartige digitale Möglichkeit für die Fachbereiche dar, die Transkripte verwenden, gleichzeitig bringen sie auch Veränderungen mit sich und erfordern deshalb eine auf den neuesten Stand gebrachte Expertise. Außerdem bringt dieser Ansatz eine neue Perspektive mit sich und das nicht nur im Bereich der politischen Sprache. Die Inhalte und die rhetorischen bzw. argumentativen Strukturen der Reden sind grundlegend für die Analyse der politischen Sprache; aber ebenso sind stimmliche Merkmale für die Wirkung einer Rede verantwortlich. Die Auseinandersetzung mit den Analyse-Tools bereichert die Forschung in diesem Bereich um ein weiteres brauchbares Instrumentarium.

2. Transkription und Segmentation von Reden

2.1. Transkriptionsmethoden: ein Exkurs

Der Begriff ‚Transkription‘ bezieht sich im Bereich der Linguistik im Allgemeinen auf zwei verschiedene Textprodukte. Zum einen versteht man unter Transkription die phonetische oder phonologische Verschriftlichung der gesprochenen Sprache unter Anwendung spezieller Alphabete, z.B. IPA oder

SAMPA. Auf der anderen Seite kann mit dem Begriff ‚Transkription‘ auch die orthographische Übertragung der gesprochenen Sprache gemeint sein. Diese findet ihre Anwendung i.d.R. in Disziplinen wie der Soziolinguistik, der Gesprächsanalyse bzw. anderen Fachgebieten, die sich vor allem mit dem Inhalt der gesprochenen Sprache beschäftigen.

Dieser Beitrag befasst sich mit der zuletzt genannten Bedeutung von Transkription, d.h. der orthographisch transkribierten gesprochenen Sprache. Die Transkription von Reden oder Gesprächen ist also ein Mittel, um z.B. Inhalte oder Interaktionen analysieren zu können. Das führt dazu, dass das Transkribieren oft als eine rein zweckmäßige und zeitaufwändige Aufgabe wahrgenommen wird, die nichts mit der eigentlichen Analyse zu tun hat, sondern nur ein Mittel zum Zweck ist. Das heißt, dass Transkripte normalerweise nicht Bestandteil der Analyse selbst sind, sondern sozusagen eine Voraussetzung, um das auditive Material sammeln und analysieren zu können. Transkripte sind aus folgenden Gründen notwendig: “necessary for investigations, for storage, reproduction, search, exemplary and presentation purposes, for it is only on the basis of a transcription that those tasks become executable” (Kirk, Andersen 2016, S. 296).

Wie erfolgt das Transkribieren zu wissenschaftlichen Zwecken üblicherweise? Da es sehr zeitintensiv ist, werden gern kurze Reden und Gespräche ausgewählt, oder es werden einfach Reden gewählt, deren Transkripte schon zur Verfügung stehen. Da jedoch die Dauer ein wichtiger Faktor dafür ist, ob eine Rede oder ein Gespräch für die Analyse ausgewählt wird, was besonders in der Politik eine wichtige Rolle spielt, bedeutet das eine Einschränkung: Es kann dazu führen, dass Reden allein aufgrund ihrer Länge von einer Analyse ausgeschlossen werden. Als weitere zeitsparende Strategie kann man sich auch auf relevante Ausschnitte einer Rede beschränken. Dabei besteht allerdings das Risiko, etwas zu übersehen, was die anschließende Analyse beeinflussen kann. Das ist insbesondere dann der Fall, wenn nur die transkribierten Teile der Rede wiedergegeben werden.

Auch bei Studien über die politische Sprache lässt sich feststellen, dass der Fokus normalerweise entweder auf schriftlichen Texten oder auf im Voraus transkribierten bzw. kurzen Reden liegt. Das Thema des Transkribierens an sich wird hier nicht unmittelbar vertieft, was die Relevanz dieses Beitrags deutlich macht. Die Untersuchungsbeispiele stammen nämlich üblicherweise von schriftsprachlichen Dokumenten oder aus Reden, für die offizielle Transkripte zur Verfügung stehen. So werden beispielsweise Wahlkampfreden, die meist vor größeren Menschenmengen gehalten werden, eher selten transkribiert. Der Grund dafür ist einerseits, dass diese Reden im Normalfall nicht im Voraus schriftlich vorliegen und andererseits, dass das Audio-Format dieser Art von Reden wegen des großen Publikums häufig Hintergrundgeräusche aufweist, die eine Transkription erschweren.

Brambilla (2007, S. 63) verweist darauf, dass es im Hinblick auf die Analyse der gesprochenen Sprache in politischen Reden eine Forschungslücke gibt, obwohl die stimmliche Umsetzung vor allem in Bezug auf die Persuasion die Hauptrolle spielt. Außerdem hebt Brambilla hervor, dass offizielle Transkripte politischer Reden oft von Stenografen geglättet wurden und deswegen keine zuverlässige Quelle für die Analyse darstellen (vgl. ebd., S. 66). Das heißt, dass die Transkriptionsprobleme nicht nur Reden betreffen, die noch nie transkribiert wurden, sondern auch diejenigen, für die offizielle Transkripte existieren. Beispielsweise stammt das offizielle Transkript einer von Brambilla analysierten 40-minütigen politischen Rede (ebd., S. 111-126) auch von Stenografen. Wenn man die Rede nämlich auf YouTube abspielt, merkt man sofort, dass die Transkription nicht mit dem gesprochenen Text übereinstimmt. Das zeigt, dass ein zuverlässigeres Transkript entstehen würde, wenn sich der Zeitaufwand reduzieren ließe und man die Transkription computergestützt durchführen könnte. Zudem zeigt es, dass Transkriptionshindernisse nicht von den Textsorten bzw. Textarten abhängig sind, sondern verschiedene Textsorten gleichermaßen betreffen.

Das Transkribieren ist nämlich ein komplexer Prozess und daher haben sich viele verschiedene Forschungsansätze dazu entwickelt, die bis heute noch keinen wirklichen gemeinsamen Nenner haben. Fuß und Karbach (2019) unterscheiden beispielsweise zahlreiche Transkriptionstypen auf der Basis der Transkriptionsregeln und der Phänomene, die verschriftlicht werden sollten (ebd., S. 20). Sie stellen fest, dass „das Transkript [...] letztendlich als das Ergebnis einer Transformation von Daten zu wissenschaftlichen Zwecken [ist]“ (ebd., S. 28). Aus diesem Grund spielt es auf jeden Fall eine Rolle, wie die Transkription gestaltet und durchgeführt wird.

Heilmann (2003) beschreibt drei verschiedene Transkriptionsmethoden und stellt dabei die Schwierigkeiten des Transkriptionsprozesses dar. Laut Heilmann (ebd., S. 142) dienen Transkriptionen dazu, „Mündlichkeit zu verschriftlichen“, wobei Parameter festgesetzt werden müssen, die eine Analyse möglich machen. In Bezug auf den Transkriptionsprozess erkennt Heilmann (ebd.) verschiedene Paradoxien des Transkribierens. Eine davon ist die „Tatsache, dass [...] die transkribierten Äußerungen in ihrer situationsgebundenen Mündlichkeit so weit als nur irgend möglich erhalten bleiben sollen“. Es scheine in der Tat widersprüchlich, die gesprochene Sprache zu verschriftlichen, wobei man gleichzeitig angehalten werde, darauf zu achten, die Kennzeichen der Mündlichkeit bei den Transkripten nicht zu verlieren. Je nach Forschungszweck der Transkripte existieren verschiedene Transkriptionsprototypen. Heilmann fokussiert sich auf drei davon (ebd., S. 144-150).

Der erste Typ ist die *integrierte Zeilenschreibung*, die laut Heilmann (ebd., S. 144) das GAT (Gesprächsanalytisches Transkriptionssystem, online

verfügbar)³ als Notationssystem verwendet. Solch ein Notationstyp benötigt keine besondere Software, vielmehr wird i.d.R. ein simples Textdokument verfasst. Obwohl Pausen und Interjektionen auch notiert werden, zielt diese Art von Transkripten darauf ab, die „textuelle Verflechtung“ (ebd.) zu verdeutlichen. Seit 2003, d.h. seit Veröffentlichung des eben zitierten Beitrags von Heilmann, hat sich das GAT-System verändert. Im Jahr 2009 wurde nämlich GAT 2 eingeführt (Selting *et al.* 2009), das der internationalen Forschungsgemeinschaft dienen soll und bei dem Transkriptionskonventionen weiterentwickelt wurden (ebd., S. 354). Heilmann (2003, S. 144) weist darauf hin, dass bei der Verwendung von GAT und seiner integrierten Zeilenschreibung aufgrund eindeutiger Konventionen Fehler reduziert und Irrtümer verhindert werden.

Die anderen beiden von Heilmann (ebd.) dargelegten Transkriptionsprototypen sind die *Partiturschreibung* und die *Zeitreihennotation*. Bei der Partiturschreibung

[werden] die unterschiedlichen Ebenen gesprochener Sprache (linguistische, paralinguistische und extralinguistische) [...] – einer Partitur vergleichbar – übereinander angeordnet, so dass eine synchronoptische Darstellung entsteht. (Heilmann 2003, S. 145)

Normalerweise wird so eine Art von Transkription von einer Software unterstützt. Das dazugehörige Notationssystem ist laut Heilmann (vgl. ebd., S. 145) HIAT (Halbinterpretatives Arbeitstranskript, online verfügbar).⁴ Während man den Umgang mit der Software erlernen muss, benötigt auch das Transkribieren relativ viel Zeit, insbesondere wenn die Transkription sehr ausführlich sein soll. Ein kostenloses häufig benutztes Programm dafür ist EXMARALDA,⁵ das heruntergeladen und ebenfalls regelmäßig aktualisiert werden muss.

Die *Zeitreihennotation* (Heilmann 2003, S. 148-149) ist schließlich ein visueller Transkriptionsprototyp, bei dem „in Analogie zur gesprochenen Sprache als Folge einzelner Laute [...] das menschliche Bewegungsverhalten als Folge von Positionszuständen einzelner Körperteile durch Ziffernfolgen abgebildet“ wird (ebd., S. 148). Diese Art von Transkription ist besonders zeitaufwändig und man braucht eine gewisse Expertise, um sie durchführen zu können. Es handelt sich eher um eine Art von Visualisierung, welche die Körpersprache einbezieht (z.B. Blickverhalten, Gestik) und mit der Verschriftlichung des gesprochenen Textes verbindet.

Heilmann (vgl. ebd., S. 150) erläutert dazu, dass die ausgewählte Transkriptionsmethode vom Untersuchungsziel abhängt. Alle drei von

³ <https://agd.ids-mannheim.de/gat.shtml> (04.06.2023).

⁴ https://www.exmaralda.org/hiat/files/azm_56.pdf (04.06.2023).

⁵ <https://exmaralda.org/en/> (25.04.2023).

Heilmann dargestellten Methoden weisen gewisse Stärken und Schwächen⁶ auf und setzen zudem eine manuelle Transkription voraus.

Das Projekt FOLK (Forschungs- und Lehrkorpus Gesprochenes Deutsch)⁷ bietet die Möglichkeit, ausgewählte transkriptionsbezogene Aspekte zu untersuchen. Allerdings stammt die gesprochene Sprache in FOLK hauptsächlich aus Interaktionen (Schmidt 2016, S. 397), weshalb ein Transkriptionsinstrumentarium ausgewählt wurde, das dafür geeignet ist. Interaktionstranskriptionen stellen weiterhin eine Herausforderung für die in diesem Beitrag präsentierten Tools dar, was in den folgenden Abschnitten genauer erläutert wird. Schmidt (ebd., S. 404) hebt vor allem den Zeitaufwand für die manuelle Transkription hervor. Er schlägt jedoch nicht vor, die automatische Spracherkennung (ASR) zu verwenden, sondern den Fokus auf die uniformierten Transkriptionskonventionen und ein benutzerfreundliches Tool zu legen, das leicht zu erlernen und zu verwenden ist (vgl. ebd., S. 404-405). Daher wurde die Software FOLKER⁸ ausgewählt, bei der der Text automatisch auf GAT-Konventionen überprüft wird und jedem/r SprecherIn eine Transkriptionspur zugewiesen werden kann (ebd., S. 407). Laut Schmidt (ebd., S. 413) kann FOLKER den Transkriptionsprozess darüber hinaus optimieren und beschleunigen, obwohl es sich immer noch um eine manuelle Transkription handelt. Im Abschnitt 3 wird auch darüber reflektiert, ob und wie die vorgestellten Tools andere Programme ergänzen können. Es wäre z.B. gut möglich, FOLKER und Web Tools zusammen zu verwenden.

Wenn auch die oben beschriebenen Methoden gültige Transkriptionsmethoden sind, sollte man auch andere Methoden kennen, um sich je nach Forschungsziel die richtige Software bzw. das richtige Tool auszusuchen.

2.2. Pausenberechnung und Pausenmarkierung

Wie schon erwähnt wurde, ist es schwierig, die genaue Dauer von Pausen wiederzugeben. Notationssysteme erlauben es normalerweise, die Dauer der Pausen anzugeben, wobei die Eingabe manuell gemacht werden muss. Dies gilt für alle oben genannten Transkriptionsprototypen.

Für die manuelle Markierung bieten Fuß und Karbach (2019, S. 44) folgende Visualisierungsmöglichkeiten. Eine Möglichkeit besteht darin, Pausen sekundengenau zu transkribieren, während eine andere Option darin besteht, die ungefähre Pausenlänge in drei definierten Längen grob zu unterscheiden. Demzufolge bezeichnet „(..)“ eine Pause von zwei Sekunden

⁶ Zum Beispiel kann die Dauer der Pausen fehlerhaft transkribiert werden, wie im folgenden Abschnitt vertieft wird.

⁷ https://agd.ids-mannheim.de/FOLK_extern.shtml (07.10.2023).

⁸ <https://agd.ids-mannheim.de/folker.shtml> (07.10.2023).

und „(...)“ eine Pause von bis zu fünf Sekunden, während längere Pausen als „(Pause)“ vermerkt werden (ebd.). FOLKER würde es ermöglichen, Pausen automatisch zu messen und zu transkribieren (Schmidt 2016, S. 408).

Bei den im Folgenden präsentierten BAS Web Tools besteht die Option, Pausen überhaupt nicht zu transkribieren oder sich für das automatische Anzeigen von Zeitmarkern zu entscheiden. Obwohl man auch hier auf Schwierigkeiten stoßen kann, ist die Fehleranfälligkeit bei der automatisierten Dauermessung der Pausen gering.

Schließlich gibt es die Variante, die Pausen überhaupt nicht zu markieren und sie direkt im Zusammenhang mit Klangtools zu beobachten und zu untersuchen. Dies reduziert nicht nur mögliche Fehler, sondern macht die Transkripte auch leichter lesbar. Mit diesem Ansatz kann man sowohl den Inhalt bzw. die textuelle Struktur als auch das Signal untersuchen.

Man könnte meinen, dass das Thema der Pausenproblematik nebensächlich sei, aber in linguistischen Disziplinen spielt es eine grundlegende Rolle. Sowohl in Interaktionen als auch in Reden sind die Pausen und ihre Dauer von großer Bedeutung. Aus diesem Grund sollten sie so präzise wie möglich berechnet und angegeben werden. Das Ergebnis von den hier präsentierten BAS Web Tools ist daher ein zeitaligniertes Transkript, mittels dessen man die verschiedenen Phänomene beobachten kann. Zudem kann man wissen, wann sie stattfinden. Die Tools verfügen dazu über verschiedene Export-Möglichkeiten in passenden Formaten. Dabei bieten sie Funktionen an, die es ermöglichen, den Wechsel zwischen Klang und Pausen zu untersuchen.

3. BAS Web Tools

BAS ist die Abkürzung für das Bayerische Archiv für Sprachsignale. Es wurde im Jahr 1995 gegründet und gehört seit 2013 zur CLARIN Infrastruktur.⁹ Das BAS umfasst verschiedene Tools für Phonetik und Sprachverarbeitung und beinhaltet auch genau die Tools, die in diesem Beitrag vorgestellt werden (vgl. Kisler *et al.* 2017). OCTRA und WebMAUS werden ausführlich präsentiert, während auf EMU-webApp nur kurz hingewiesen wird.

Die BAS Web Tools können entweder auf dem Browser benutzt werden (sie sind kostenfrei zugänglich mittels einer E-Mail-Adresse, die der CLARIN-Infrastruktur zugehörig ist) oder direkt in der Programmiersprache R¹⁰ durch das Paket emuR¹¹ aufgerufen werden (Poerner, Winkelmann 2018, S. 141): „the BAS-emuR interface, which allows users to apply BAS web services to

⁹ <https://www.bas.uni-muenchen.de/Bas/BasHomedeu.html> (25.04.2023).

¹⁰ R ist eine Programmiersprache für statistische Berechnungen und Grafiken: <https://www.r-project.org/> (09.02.2024).

¹¹ R-Pakete sind Erweiterungen für die Programmiersprache R. Sie enthalten Codes, Befehle und Daten. emuR: <https://cran.r-project.org/web/packages/emuR/index.html> (09.02.2024).

their EMU databases from within the R environment“. Laut Poerner und Winkelmann (ebd., S. 143) liegt der größte Vorteil dieser Option darin, dass man direkt „[a] ready-made EMU database“ bekommt, was heißt, dass man keine Formate konvertieren muss (ebd.). Ein Nachteil, der genannt wird, ist, dass man eine stabile Internetverbindung braucht. Des Weiteren müssen die Dateien temporär auf dem BAS gespeichert werden (ebd.). Man muss also die Programmiersprache R und das Paket emuR gut beherrschen, um diese Anwendung nutzen zu können. Wenn man die große Palette an BAS Tools zum ersten Mal verwenden möchte, ist es unwahrscheinlich, dass man emuR schon kennt. Aus diesem Grund wurde in diesem Beitrag beschlossen, die Tools einzeln zu besprechen, wenn sie auch mittels emuR auf R aufgerufen werden könnten. Dies ist u.a. deshalb sinnvoll, weil diese Tools auch von LinguistInnen aus Bereichen außerhalb der Phonetik verwendet werden können.

BAS Web Tools unterscheiden sich von Transkriptionsprogrammen, die als Software entwickelt wurden, d.h. die auf dem Betriebssystem eines Geräts heruntergeladen werden und nicht direkt auf einem Server laufen, wie z.B. EXMARaLDA und ELAN.¹² Die unterschiedlichen Leistungen der Betriebssysteme beeinflussen nämlich auch die Leistung der Software. Die webbasierten Tools sind hingegen von der Leistung des jeweiligen Betriebssystems unabhängig, weil die Verarbeitung der Dateien auf dem Server erfolgt (vgl. Kisler *et al.* 2016, S. 3882), d.h. dass man nur ein Gerät und eine stabile Internetverbindung dazu braucht. Die Einstellungen der verwendeten Sprachen können leicht geändert werden, da man sie einfach online auswählen kann, ohne dass die Installierungssprache mit der jeweiligen Software interferiert (vgl. ebd.). Die webbasierten Tools wurden gerade deshalb entwickelt, um Schwierigkeiten bei der Verwendung lokaler Software zu umgehen. Das bedeutet, dass softwarebedingte Probleme umgangen bzw. ausgeschlossen werden können.

Gelegentlich kann es beim Hochladen großer Dateien auch passieren, dass die Software überfordert ist und es sogar zu einem Systemabsturz kommt. Dies ist bei den webbasierten Tools nicht zu erwarten:

This allows access for multiple users at the same time without slowing down the processing for other users. All web interfaces that allow to upload multiple files per drag and drop make use of this feature and process, for one user, multiple files at the same time. (Kisler *et al.* 2016, S. 3882)

Mit BAS webbasierten Tools ist es möglich, eine Verknüpfung unterschiedlicher Programme, die nicht webbasiert sind, herzustellen (siehe Abschnitt 2.1). Wie Abbildung 1 zeigt, kann man beispielsweise die (externe)

¹² <https://archive.mpi.nl/tla/elan> (25.04.2023).

Software EXMARaLDA mit WebMAUS, das zu den BAS Web Tools gehört, verbinden.

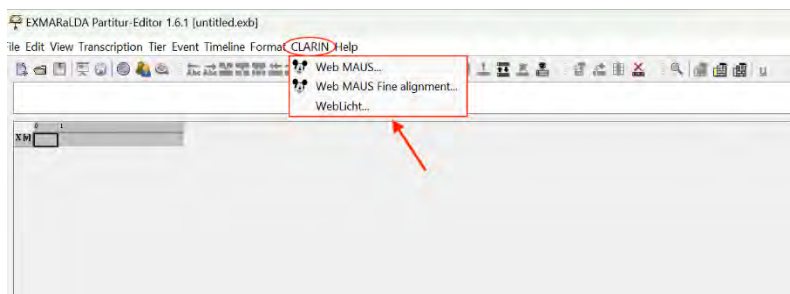


Abbildung 1

Zugang zu BAS Web Services durch EXMARaLDA (siehe Markierung).

Auch weitere web- und softwarebasierte Programme können durch den Export in beliebig viele Dateiformate verwandelt werden.

Außerdem sind die BAS Web Tools benutzerfreundlich und wurden im Laufe der Zeit für zahlreiche nicht phonetische Disziplinen (z.B. *Oral History*) weiterentwickelt (vgl. Pömp, Draxler 2017, S. 145). Somit stellt sich auch die Frage, ob die Tools, die für phonetische Fragestellungen und Sprachverarbeitung entwickelt wurden, auch für andere sprachwissenschaftliche Disziplinen wie die Politolinguistik geeignet sind. Im Folgenden werden zwei ausgewählte Tools vorgestellt: OCTRA¹³ und WebMAUS General¹⁴.

3.1. OCTRA

OCTRA (*OrthographiC TRAnscription*) ist ein Tool, das die orthographische Transkription von Audio Files ermöglicht. Es wird einfach mit dem Browser verwendet, d.h. dass man keine Software herunterladen muss. Außerdem bietet OCTRA zahlreiche verschiedene Output-Formate, damit man die Transkripte mit der jeweils gewählten Software und anderen Tools weiter bearbeiten kann. Das Handbuch für OCTRA¹⁵ steht online zur Verfügung, so dass man sich einerseits über die Funktionen und andererseits über die Ziele dieses Tools informieren kann. Der Ablauf des Transkribierens mit OCTRA wird in einzelnen Schritten anhand von Beispielen und Bildern erläutert.

Was die Konventionen für die Transkription angeht, braucht man normalerweise einige Regeln. Bei OCTRA existieren allerdings keine vorgegebenen Regeln wie bei GAT, vielmehr können die transkribierenden Personen ihre eigenen Richtlinien definieren oder entsprechende Icons

¹³ <https://clarin.phonetik.uni-muenchen.de/apps/octra/octra/login> (16.05.2023).

¹⁴ <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSGeneral> (25.04.2023).

¹⁵ <https://clarin.phonetik.uni-muenchen.de/apps/octra/manual/1.4.0/> (25.04.2023).

auswählen, wie Abbildung 2 zeigt. Dies funktioniert, solange die Konventionen in den Transkripten konsistent sind.

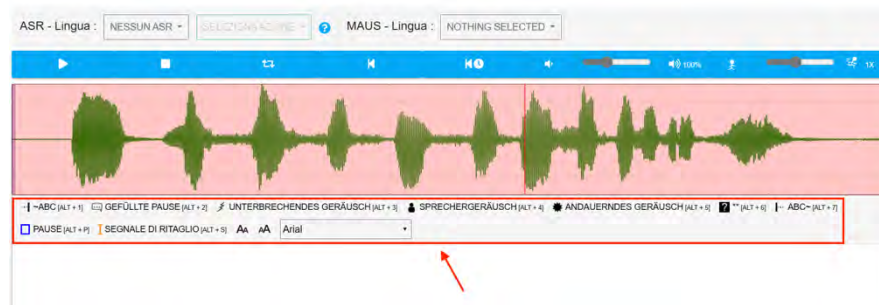


Abbildung 2
Icons auf OCTRA-Transkriptionskonsole (siehe Markierung).

Was den Prozess der Transkriptionen bei OCTRA angeht, merken Pömp und Draxler (2017, S. 146) Folgendes an:

OCTRA (for *OrthographiC TRAnscription*) is based on the concept of time-aligned transcription chunks: transcribers insert segment boundaries at appropriate positions in the signal, usually within pauses or at the end of intonation phrase.

In OCTRA schreibt man unabhängig von der Anzahl der SprecherInnen auf einer einzigen Spur. Die verschiedenen InterlokutorInnen können unterschieden werden, indem man ein Zeichen setzt, wenn eine andere Person zu sprechen beginnt. Üblich ist die Kennzeichnung der einzelnen SprecherInnen mit <S0,1,..>. Es gibt zwar die Möglichkeit, weitere Spuren hinzuzufügen, dies ist allerdings nur im lokalen Modus möglich (siehe Abbildung 3). Man sollte beachten, dass es nicht möglich ist, die verschiedenen Spuren untereinander und somit gleichzeitig zu betrachten. Wenn man Gespräche visualisieren möchte, dann sollte man Konventionen einsetzen, die es ermöglichen, beispielsweise Überlappungen und andere typische Elemente auf nur einer Spur zu zeigen. So ist eine Möglichkeit, Überlappungen mit dem Zeichen <ovl> zu markieren. Kirk & Andersen (2016, S. 296) verweisen in diesem Zusammenhang auf den organisatorischen Vorteil eines aus mehreren Spuren bestehenden Transkripts, was die Übersichtlichkeit, Klarheit und Lesbarkeit angeht. Das heißt, dass man sich bei der Entscheidung für das passende Programm von den Forschungszwecken seines Projekts leiten lassen soll. Dabei sollten sowohl technologische als auch methodologische Aspekte berücksichtigt werden (Schmidt 2016, S. 401).

Wenn man OCTRA öffnet, erscheinen zwei Modi (Abbildung 3): der Online-Modus (links) und der lokale Modus (rechts). Der Online-Modus eignet sich ein Projekt, das nur für seine Mitglieder zugänglich ist. Beim lokalen

Modus kann man autonom Audio Files hochladen und die Transkription anfangen.

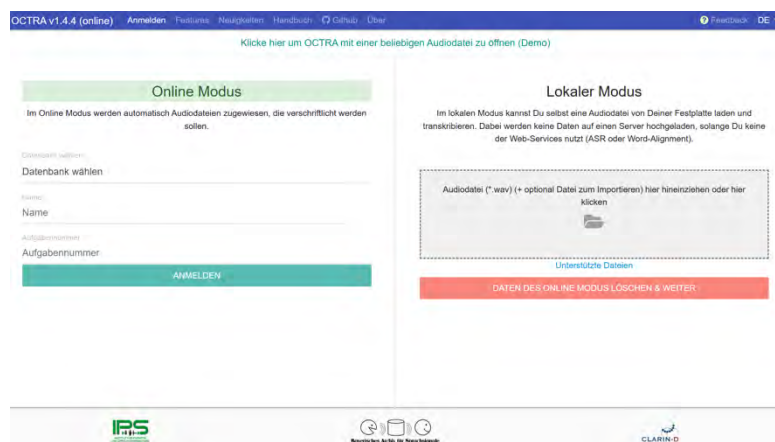


Abbildung 3
Online-Modus (links) und lokaler Modus (rechts) auf OCTRA.

Transkriptionseinheiten können unterschiedlich lang sein und sie werden beliebig aufgeteilt. Damit der Redefluss nicht unterbrochen wird, werden die Grenzmarker üblicherweise bei Pausen platziert (Abbildung 4).



Abbildung 4
Trennungszeichenplatzierung auf OCTRA (siehe Markierung).

Danach kann man auf die jeweilige Einheit klicken und ein Segment nach dem anderen transkribieren (Abbildung 5).

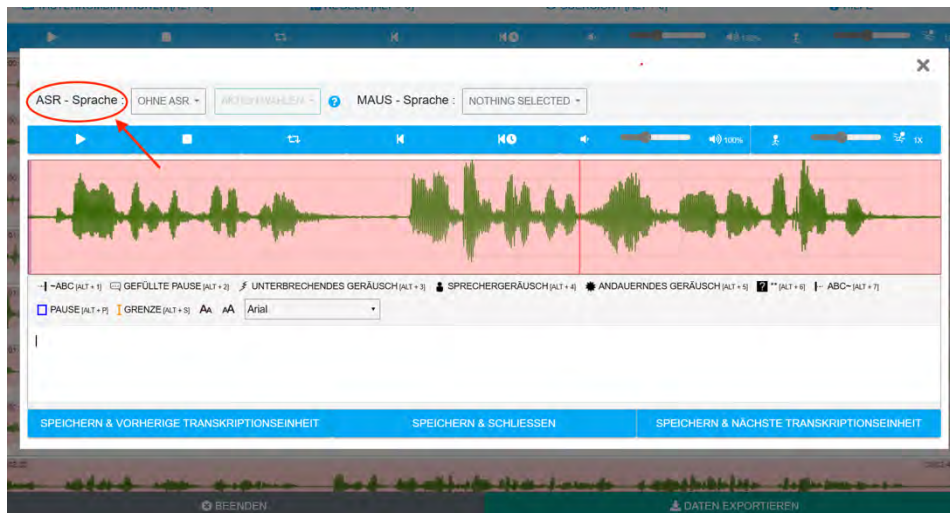


Abbildung 5

Überblick über ein einzelnes zu transkribierendes Segment und die ASR in OCTRA.

In der Abbildung 5 oben links gibt es die Möglichkeit, die automatische Spracherkennung (ASR) zu nutzen und sie entweder nur auf die geöffnete Einheit oder auf den ganzen Text anzuwenden. Die Anwendung der automatischen Spracherkennung wird im Abschnitt 4 vertieft.

Nach Beendigung der Transkription kann man diese mit einem Klick auf „Übersicht“ anschauen. Das Transkript ist schließlich in verschiedenen Formaten exportierbar, u.a. im txt-Format, das beispielsweise für WebMAUS benötigt wird (siehe Abbildung 6).

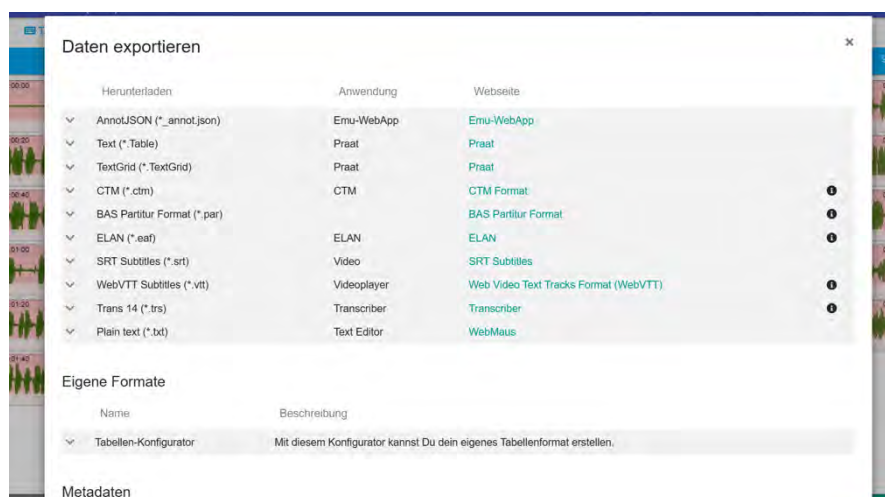


Abbildung 6

Tabelle der Output-Formate in OCTRA.

Die auf diese Weise erstellten Texte enthalten die orthographische Transkription und können für alle möglichen linguistischen Analysen verwendet werden. Innerhalb des txt-Formates können verschiedene Exportvarianten ausgewählt werden. Beispielsweise kann man entscheiden, ob

die Zeitangaben in den Text integriert werden sollen, oder ob der Text ohne Zeitangaben gespeichert werden soll. Beispiele von Transkripten mit und ohne Zeitangaben aus den italienischen und deutschen politischen Reden sind im Abschnitt 4 zu finden.

3.2. WebMAUS

WebMAUS (*Munich AUtomatic Segmentation*) ist ein weiteres BAS Tool und wird oft im Zusammenhang mit OCTRA benutzt. Dieses Tool ist hilfreich für unterschiedliche Anwendungen, z.B. für Redeanalysen auch im Bereich der (Polito)Linguistik, weil es, wie bereits erwähnt, die Möglichkeit bietet, Signale zu analysieren. Wie im Abschnitt 4 dargelegt wird, können zunächst OCTRA und anschließend WebMAUS eingesetzt werden. WebMAUS bietet die automatisierte Segmentierung bzw. Alignierung von Audiosignalen und den jeweils zugehörigen Transkripten¹⁶ nicht nur für die Segmente, die mit OCTRA bearbeitet wurden, sondern auch für einzelne Laute: Es können also sowohl die Position jedes Wortes im Signal bestimmt als auch die Pausen präzise gemessen werden. Um eine Analyse durchzuführen, die auf der einen Seite das Signal und auf der anderen textuelle Bestandteile berücksichtigt, ist WebMAUS bestens geeignet. Außerdem müssen Pausen in den Transkripten, wie bereits erwähnt, nicht unbedingt geschrieben werden, da man sie in WebMAUS direkt visualisieren kann. Insbesondere kann man je nach Forschungszweck verschiedene Output-Transkripte einrichten: Wenn man z.B. lediglich die Wortwahl oder die Syntax eines politischen Textes analysieren will, spielen die Pausen eine untergeordnete Rolle. Ganz davon abgesehen, dass die Transkripte ohne Pausenmarkierung leichter zu lesen sind. Wenn man dagegen den Redefluss untersuchen will, kann man dies mit den mit Hilfe von WebMAUS verarbeiteten Files im Detail untersuchen, da der Text mit der Sprachaufnahme aligniert wird.

Grundsätzlich gibt es drei Varianten von WebMAUS, die zur Verfügung stehen: WebMAUS Basic,¹⁷ WebMAUS General¹⁸ und Pipeline with(out) ASR.¹⁹ WebMAUS Basic (Abbildung 7) ist, wie der Name vermuten lässt, die Basisversion von WebMAUS und enthält weniger Funktionen als die erweiterte Version WebMAUS General.

¹⁶ <https://www.clarin.eu/showcase/webmaus-automatic-segmentation-and-labelling-speech-signals-over-web> (11.10.2023).

¹⁷ <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic> (25.04.2023).

¹⁸ <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSGeneral> (25.04.2023).

¹⁹ <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/Pipeline> (25.04.2023).



Abbildung 7

Homepage von WebMAUS Basic und Überblick über die Funktionen (siehe Markierung).

Abbildung 7 zeigt die Startseite der Browserversion von WebMAUS Basic, auf der bereits ersichtlich ist, dass nur wenige Einstellungen zur Verfügung stehen. Man muss jeweils eine Audiodatei im wav-Format und eine gleichnamige Text-Datei im txt-Format hochladen, um das Programm zu starten. Daneben kann man die gewünschte Sprache und ein passendes Output-Format auswählen. Im unteren Feld werden hingegen Status- sowie Fehlermeldungen angezeigt. Nachdem die Dateien verarbeitet und exportiert wurden, kann man sie herunterladen. Für größere Dateien wird ein komprimierter ZIP-Ordner automatisch erstellt. Leider besteht in WebMAUS Basic nicht die Möglichkeit, sich per E-Mail über den Status der Verarbeitung benachrichtigen zu lassen (dies gehört zu den *Expert Options*), selbst wenn die Anzahl der zu bearbeitenden Dateien groß ist und der Export lange dauert.

Das letztgenannte Problem wurde von WebMAUS General gelöst. In dieser fortgeschrittenen Version von WebMAUS Basic gibt es eine Reihe erweiterter Einstellungen (*Expert Options*), wie die Abbildungen 8-11 darstellen:



Abbildung 8

Service Options von WebMAUS General.

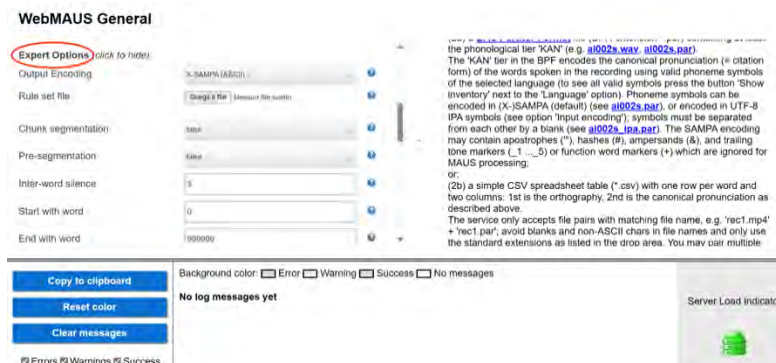


Abbildung 9
Expert Options von WebMAUS General (siehe Markierung).



Abbildung 10
Weitere Expert Options von WebMAUS General.

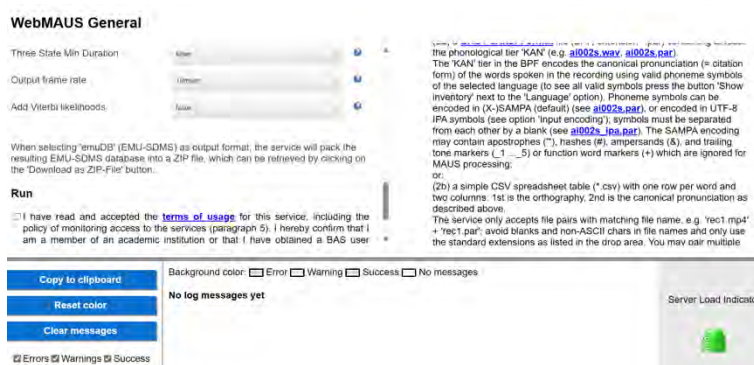


Abbildung 11
Weitere Expert Options von WebMAUS General.

Unter diesen Optionen gibt es auch die Möglichkeit, sich Benachrichtigungen per E-Mail senden zu lassen. Obwohl WebMAUS General die erweiterte Version von WebMAUS Basic ist und daher zahlreiche zusätzliche Funktionen umfasst, bietet es keine Lösung für ein wichtiges Problem an, dass nämlich WebMAUS General nur Dateien im BAS Partitur Format (BPF) erkennt. Das heißt, dass diese jedes Mal zuerst konvertiert werden müssen. Zwar wurde das

BAS-Tool Annotconv²⁰ extra dafür entwickelt, aber gleichzeitig kann man damit nur BPF in andere Formate konvertieren und nicht umgekehrt. Die Verwendung von WebMAUS General könnte aufgrund dieser Formatprobleme aufwendig sein.

Um WebMAUS General dennoch nutzen zu können, ohne die Dateien vorher mittels BPF konvertieren zu müssen, wird von BAS die Anwendung Pipeline with(out) ASR als mögliche Lösung angeboten. Die Pipeline ruft WebMAUS General auf und ermöglicht auch dort, die zahlreichen erweiterten Einstellungen zu öffnen. Anschließend müssen dann, wie auch in WebMAUS, jeweils eine wav- und eine txt-Datei hochgeladen werden. Auch hier muss man wieder darauf achten, dass zusammengehörige Dateipaare denselben Namen haben. Die Abbildung 12 zeigt die Startseite der Anwendung Pipeline without ASR:



Abbildung 12
Homepage von Pipeline without ASR.

Mit Hilfe von Pipeline kann man die von WebMAUS General bearbeiteten Dateien in einem beliebigen Output-Format erhalten und sie anschließend mit weiteren Tools oder weiteren Softwareprogrammen analysieren. Das Programm verfügt über zahlreiche Output-Formate und verfügbare Sprachen, wodurch die Weiterverarbeitung mit weiteren Anwendungen erleichtert wird. Für das erstellte Korpus werden sowohl TextGrid als auch annotJSON als Output-Formate benutzt, weitere Formate die zu anderen Applikationen passen, wie z.B. ELAN, können gewählt werden. Des Weiteren kann die Zahl der Spuren, die man mit WebMAUS bearbeiten möchte, festgelegt werden. Es wurde bereits angemerkt, dass die Pausen auf OCTRA nicht unbedingt annotiert werden müssen, da WebMAUS sie bearbeitet. Das Signal der Reden kann dann zusätzlich beispielweise auf Praat²¹ oder auf EMU-webApp²² angezeigt und analysiert werden (Abschnitt 4).

²⁰ <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/AnnotConv> (25. 04. 2023).

²¹ <https://www.fon.hum.uva.nl/praat/> (12.10.2023).

²² <https://ips-lmu.github.io/EMU-webApp/> (25.04.2023).

3.3. Potenzial und Herausforderungen der Verwendung von BAS Web Services

Die BAS Web Tools bieten diverse Verwendungsmöglichkeiten in unterschiedlichen Disziplinen, so auch in der (Polito)Linguistik. Auf einige Vorteile und potenzielle Probleme bzw. Hürden soll abschließend zusammenfassend eingegangen werden.

3.3.1. Potenzielle Vorteile von BAS Web Tools

Aus der Sicht der BenutzerInnen ist einer der wichtigsten Vorteile zweifellos, dass man kein Programm auf dem Rechner installieren muss. Installierte Programme sind im Allgemeinen oft problematisch, da es nicht immer möglich ist, die neueste und gleichzeitig die für das eigene Betriebssystem kompatible Version herunterzuladen. Diese Problematik erscheint im ersten Moment oft unwichtig, doch wenn man für ein Forschungsprojekt Dateien mit geeigneter Software transkribieren und analysieren möchte und dafür viel Zeit und Mühe investiert, ist es relevant, das Risiko einer niedrigen Leistung des Betriebssystems oder des Systemabsturzes auf jeden Fall zu vermeiden.

Ein weiterer Vorteil ist, dass BAS Web Tools benutzerfreundlich sind. BenutzerInnen können zahlreiche Erklärungen auf der Homepage, Handbücher und Online-Tutorials konsultieren. Da man die Verwendungssprache auswählen kann, sind die Status- und Fehlermeldungen im Benachrichtigungsfeld zudem für alle verständlich.

Dazu kommt ein weiterer Pluspunkt der BAS Web Tools, und zwar die Interoperabilität sowohl mit anderen webbasierten als auch mit nicht webbasierten Programmen. Aufgrund der Vielfalt möglicher Output-Formate kann man die Tools in Verbindung mit anderen Instrumenten oder Softwareprogrammen nutzen, insbesondere da man einige Applikationen nur mit bestimmten Formaten verarbeiten kann (u.a. Praat mit TextGrid). Dies ist für die Arbeit in Forschungsprojekten, in denen es üblich ist, mit einer Vielzahl verschiedener Programme zu arbeiten, von großer Bedeutung.

Letztendlich wird es dank OCTRA möglich, den Transkriptionsprozess in hohem Maß zu beschleunigen, insbesondere mit der ASR, die für die Transkription, auch im Bereich der politischen Sprache, eine entscheidende Rolle spielen kann (siehe Abschnitt 4). Bisher sind keine Studien darüber in der Politolinguistik bekannt, daneben fehlen auch Beispielfälle, wo die ASR-Systeme dieses Beitrags für das Transkribieren politischer Reden verwendet werden. Einerseits hat dies auf jeden Fall mit der schon dargelegten traditionellen Trennung zwischen textuellen Elementen und dem Signal zu tun, das in der Politolinguistik selten berücksichtigt wird. Andererseits ist das Potenzial dieses Verfahrens in der Politolinguistik meistens noch nicht

genügend erkannt worden, so dass dieser Beitrag dafür gedacht ist, auf diese Verwendungsmöglichkeiten hinzuweisen.

3.3.2. *Hürden bei Bas Web Tools*

Probleme bei der potenziellen Verwendung von BAS Web Tools können verschiedener Art sein. Die erste, vielleicht triviale Problematik ist, dass man eine stabile Internetverbindung braucht. Üblicherweise ist eine Internetverbindung mittels WLAN vielerorts verfügbar, dennoch ist diese nicht immer und überall stabil. Des Weiteren benötigt man eine E-Mail-Adresse, die mit einer Universität oder einem Forschungszentrum der CLARIN-Infrastruktur verbunden ist, um kostenlosen Zugang zu Webdiensten und Inhalten des Repository zu erhalten.

Der Umgang mit technischen Schwierigkeiten könnte eine weitere Herausforderung darstellen. Man kann zwar jederzeit den Support um Hilfe bitten und ein Feedback hinterlassen, doch selbst wenn man schnell eine Antwort bekommt, ist sie meist nicht unmittelbar.

In Bezug auf die Transkription mit OCTRA ist einer der Nachteile, dass man lediglich auf einer einzelnen Spur transkribieren kann. Wenn mehrere SprecherInnen interagieren, wäre es hilfreich, verschiedene Personen auf unterschiedlichen Spuren gleichzeitig transkribieren zu können. Immerhin verfügt OCTRA im lokalen Modus über die Option, mehrere Spuren zu benutzen, allerdings nicht gleichzeitig. Daher ist das Tool momentan nicht besonders geeignet, Interaktionen mit vielen Überlappungen und Sprecherwechseln zu transkribieren.

Das letzte Problem, das hier genannt werden soll, bezieht sich auf die Anwender-Reichweite der Tools. Die BAS Web Tools sind nicht überall bekannt. An italienischen Universitäten tendiert man z.B. dazu, lokal installierte Programme wie EXMARaLDA oder ELAN zu verwenden. Das kann selbstverständlich an eigenen Präferenzen und Forschungsbedürfnissen liegen, doch fehlt der Vergleich mit Alternativen. Der Grund, warum die Tools in verschiedenen linguistischen Disziplinen nicht so bekannt sind, ist vermutlich auch, dass sie nur in Verbindung mit der Phonetik wahrgenommen werden. Die zuletzt genannte Tatsache kann auch die geringe Reichweite in der (Polito)Linguistik erklären. Es scheint also eine Frage der Zeit zu sein, bis die Tools auch in Bereichen außerhalb der Phonetik bekannter werden. Einige der Tools sind zwar eher für die rein phonetische Analyse gedacht, doch andere, wie die hier vorgestellten OCTRA und WebMAUS, können auch in anderen linguistischen Disziplinen genutzt werden und den Prozess des Transkribierens und der Analyse erheblich erleichtern.

4. Web Tools in der Politolinguistik am Beispiel italienischer und deutscher politischer Reden

In diesem Abschnitt wird eine praktische Anwendung der präsentierten Tools gezeigt. Die exemplarischen Reden, die dabei Verwendung finden, stammen aus dem Korpus des genannten Promotionsprojekts.

Das Korpus besteht aus 22 italienischen und 22 deutschen politischen Reden, die zwischen 2015 und 2022 gehalten wurden. Die Hälfte der Reden stammt von Männern und die andere Hälfte von Frauen, wobei auf die Ähnlichkeit in der politischen Orientierung geachtet wurde. Die Gesamtdauer der italienischen Reden beträgt ca. 6:48 Stunden, die der deutschen Reden ca. 5:58 Stunden (nach Draxler 2023, S. 20). Im Doktorat geht es um Genderunterschiede in politischen Reden in Italien und in Deutschland und das Korpus wurde im Hinblick darauf ausgewählt. Die 44 untersuchten Reden wurden aufgrund unterschiedlicher Anlässe gehalten, bei einigen Reden handelt es sich um Wahlkampfreden, während andere Parlamentsreden sind. Für das Vorhaben dieses Beitrags sowie des Doktorats war es nicht relevant, dass die Reden in derselben Situation stattfanden bzw. zur gleichen Textsorte gehörten, weil sich der Prozess der Transkription nicht ändern würde. Die ausgewählten Reden dienen hier als Beispiel, um verschiedene Web Tools zu präsentieren, Genre und Textsorte stellen in diesem Beitrag keine bedeutende Rolle dar. Die Tools werden vorgestellt, um SprachwissenschaftlerInnen zu unterstützen, die für ihre Untersuchungen Transkripte von (politischen) Reden benötigen.

Die Reden wurden auf der Videoplattform YouTube gesammelt und zunächst durch die kostenlose Software Audacity²³ konvertiert: „the audio from the original videos [gets] converted to wav mono audio files with a 16 bit linear quantization and 16 kHz sample rate“ (Draxler ebd., S. 19). Bei der Namensgebung der Dateien ist in der Regel darauf zu achten, dass die Tools Leerzeichen und Sonderzeichen häufig nicht erkennen. Darüber hinaus ist es sinnvoll, die Dateien mit eindeutig zuzuordnenden Bezeichnungen zu benennen. Dadurch kann man die Reden schnell wiedererkennen und einheitliche Bezeichnungen generieren. Außerdem muss auch beachtet werden, dass die Transkriptions- und Audiofiles der jeweiligen Rede genau denselben Namen haben sollten, sonst werden sie z.B. von WebMAUS nicht erkannt. Im Fall des Korpus wurden die Reden mit Kürzeln aus Namen, Sprache, Geschlecht der SprecherInnen, Partei und Datum der Rede kodiert (ebd., S. 20).

Da YouTube-Videos meist nicht professionell aufgenommen sind, stellen Geräusche, unverständliche Sätze, Unterbrechungen, usw. besondere Herausforderungen dar, was mit einer rein manuellen Transkription kaum zu

²³ <https://www.audacityteam.org/> (04.02.2024).

bewältigen ist. Die hier vorgestellten Web Tools sind daher eine hilfreiche Unterstützung.

Im Fall des vorliegenden Korpus wurden die Reden zunächst automatisch mit Systemen der ASR (Automatic Speech Recognition) transkribiert. Für die italienischen Reden wurde die ASR der Fondazione Bruno Kessler in Trient angewendet, während die deutschen Reden mit dem IBM Watson ASR Service des BAS transkribiert wurden (Draxler 2023, S. 20). Wie Draxler (vgl. ebd.) erwähnt, kann man den IBM Watson ASR Service direkt in OCTRA aufrufen (siehe Abbildung 5, Abschnitt 3.1). Dadurch kann man auf dem Interface von OCTRA die automatischen Transkripte anhand der ASR sehen und sie im Folgenden mit OCTRA korrigieren.

Wissenschaftlich gibt es verschiedene Parameter, um die Effizienz der Transkription messen zu können, so dass man die Leistung mit und ohne ASR vergleichen kann. Draxler (2023, S. 17) benennt in seiner Pilotstudie folgende Faktoren: *raw real-time factor*, *adjusted real-time factor* und *activity factor*.²⁴ Bei Draxler (ebd.) geht es um technische Aspekte bzw. Ergebnisse und die Studie ist noch vorläufig. Hingegen wird sich dieser Beitrag nicht mit diesen Elementen beschäftigen und daher wird er keine technischen Vergleiche darstellen. Darüber hinaus wurde solch ein Verfahren für Korpora der Politolinguistik bisher nie eingesetzt und diese Methodik wurde das erste Mal im erwähnten Promotionsprojekt verwendet.

Fuß und Karbach (2019, S. 93) verweisen auf die technischen Mängel der Spracherkennungssoftware, die üblicherweise verwendet wird, um Diktate zu erstellen. Diese Programme sind nicht leistungsfähig genug, wenn die Aufnahmequalität schlecht ist oder wenn es Nebengeräusche gibt. Die Autoren (ebd., S. 93-94) legen darüber hinaus dar, dass die Spracherkennungsprogramme es nicht erlauben, Transkriptionszeichen während des Transkriptionsprozess anzubringen. Diese Zeichen können erst am Ende hinzugefügt werden, so dass der Transkriptionsprozess wiederum aufwendig wird. Des Weiteren kann man davon ausgehen, dass die von Fuß und Karbach erwähnten Programme wahrscheinlich keine Möglichkeit bieten, das Audio bzw. Teile des Transkripts zu segmentieren. Außerdem weisen die Autoren in Bezug auf die Spracherkennung mittels künstlicher Intelligenz (ebd., S. 94-95) nur kurz darauf hin, dass noch kein Vergleich mit der manuellen Transkription angestellt wurde. Hier wird auf die künstliche Intelligenz fürs Transkribieren nicht eingegangen, wobei sie sicherlich der ideale nächste Schritt für die Transkription in der (Polito)Linguistik ist.

Dass die ASR eine vielversprechende Ressource sein kann, ist auf jeden Fall mindestens ein Erfahrungswert. Das benannte Forschungsprojekt hat tatsächlich nicht nur den Anspruch, eine politolinguistische Untersuchung, sondern auch eine Lösung für bisherige Schwierigkeiten bei der Transkription

²⁴ Kursive Markierungen im Original.

zu bieten. Anlässlich der Masterarbeit, die dem Promotionsprojekt vorausgegangen ist, wurde ein Korpus von Wahlkampfreden erstellt, das wegen der Grenzen der manuellen Transkription (in dem Fall mit EXMARaLDA) und des daraus folgenden Zeitaufwands relativ klein war. BAS Web Tools bieten eine alternative Methodik auch für die politische Sprachanalyse, die vielleicht noch nicht mit technischen Details dargestellt werden kann, aber deren Potenzial anfängt, von Studien untersucht zu werden. Da bereits ein Ergebnis aus der vorhergegangenen Erfahrung vorliegt, konnte das Korpus der Doktorarbeit mehr als fünfmal größer sein. Außerdem wurde für die Erstellung des Korpus nur ungefähr ein Drittel der Zeit benötigt. Aus diesen Gründen kann die Verwendung von OCTRA und WebMAUS eine mögliche Alternative darstellen, die sich zwar noch nicht auf bewiesene technische Daten berufen kann, aber zweifelsohne für WissenschaftlerInnen, die Transkripte von langen (politischen) Reden mit Hintergrundgeräuschen benötigen, interessant sein kann.

Für das betrachtete Korpus wurde OCTRA im Online-Modus genutzt und ein Projekt für die Reden erstellt. OCTRA wurde in diesem Fall auch für die Korrekturen von bereits transkribierten Reden verwendet, nachdem die ASR eingesetzt wurde. Im Folgenden sind zwei Beispiele von Transkripten aufgeführt, jeweils mit und ohne Zeitangaben (siehe Abschnitt 3.1), die aus OCTRA exportiert wurden:

<ts="00:03:19.134"/> als der kleine Junge in Frankfurt mutwillig vor den Zug
gestoßen wurde
<ts="00:03:22.963"/> traten Sie schweigend Ihren Urlaub an in einem gut
regierten Land
<ts="00:03:27.384"/> hätte das Staatsoberhaupt Trauerbeflagung angeordnet
aber es erfolgte nichts (de_chru_m_afd_08-11-2019)

als der kleine Junge in Frankfurt mutwillig vor den Zug gestoßen wurde traten
Sie schweigend Ihren Urlaub an in einem gut regierten Land hätte das
Staatsoberhaupt Trauerbeflagung angeordnet aber es erfolgte nichts
(de_chru_m_afd_08-11-2019)

<ts="00:04:49.650"/> anche come tutti gli altri lavoratori
<ts="00:04:53.130"/> e poi per carità non faccio il giudice non faccio il
magistrato non faccio l'avvocato però alcune sentenze ti lasciano senza parole
(it_salv_m_leg_01-05-2019)

anche come tutti gli altri lavoratori e poi per carità non faccio il giudice non
faccio il magistrato non faccio l'avvocato però alcune sentenze ti lasciano senza
parole (it_salv_m_leg_01-05-2019)

WebMAUS wurde mit Pipeline without ASR (Abschnitt 3.2) eingesetzt und vier Spuren wurden bearbeitet. Diese können in Abbildungen 13-14 auf Praat und EMU-webApp visualisiert werden. In der Abbildung 13 sind das Signal und die mit WebMAUS bearbeiteten Spuren zu sehen:

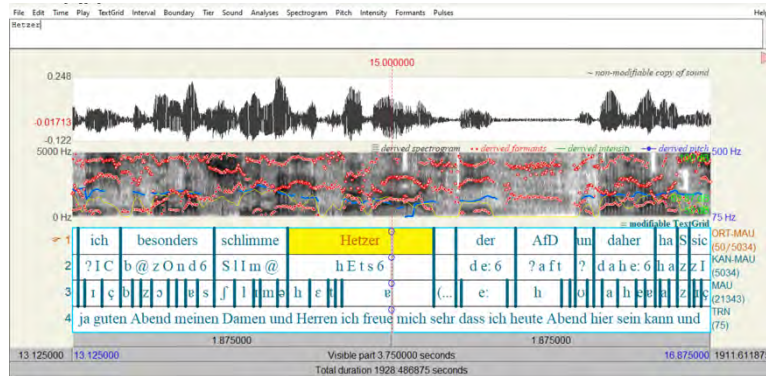


Abbildung 13

Die mit WebMaus alignierten Spuren, die auf Praat geöffnet und analysiert werden können.

Wenn die Transkription mit OCTRA durchgeführt und mit WebMAUS segmentiert bzw. aligniert wird, kann das in einem komplexen Signal mit den diversen Spuren auf EMU-webApp angezeigt werden (Abbildung 14).

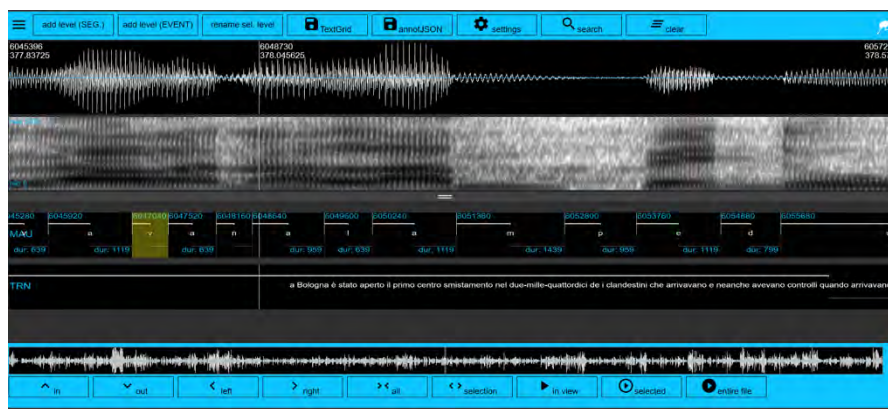


Abbildung 14

Signal und Transkriptionsspuren auf EMU-webApp.

Wenn EMU-webApp direkt im Browser verwendet und nicht über emuR aufgerufen wird, werden nicht alle vier Spuren angezeigt. Dies ist zweifelsohne ein Aspekt, der durch weitere Studien und Entwicklungen verbessert werden kann. Wie bereits erwähnt, wird R in sprachwissenschaftlichen Disziplinen selten benutzt und es wäre wünschenswert, EMU-webApp auch direkt im Browser nutzen zu können, ohne dass die Visualisierung der mit WebMAUS erstellten Spuren beeinträchtigt wird.

5. Fazit

Es gibt zur Zeit zahlreiche Möglichkeiten, die orthographische Transkription mittels verschiedener Software und Tools durchzuführen, und die in diesem Beitrag näher beleuchteten Anwendungen bieten eine zwar noch wenig bekannte, aber umso spannendere und nützliche webbasierte Alternative zu bisher genutzten Softwareprogrammen. Im Beitrag wurden die Vor- und Nachteile, sowie die Unterschiede zu herkömmlichen Methoden erörtert, um aufzuzeigen, inwiefern sie auch in anderen Disziplinen und Bereichen der Linguistik, wie der (Polito)Linguistik, Anwendung finden könnten.

Üblicherweise liegt der Analysefokus in der Linguistik entweder auf sprachwissenschaftlichen Eigenschaften oder auf dem Signal. Es gibt meist keinen Kompromiss bzw. Verknüpfungen sprachwissenschaftlicher Interessensschwerpunkte. Nichtsdestotrotz sind diese zwei Ebenen vor allem in der politischen Sprache eng miteinander verbunden: Wie man sich äußert, trägt zur Wirksamkeit der Inhalte der Reden bei und obwohl die politischen Programme sowie oft auch einige Reden schriftlich zur Verfügung stehen, besteht die politische Kommunikation vor allem aus der gesprochenen Sprache. Aus diesem Grund wird eine Analyse, die diese beiden Ebenen verknüpft, zahlreiche Aspekte vertiefen.

Die präsentierten Tools könnten an die verschiedenen Forschungszwecke in der (Polito)Linguistik angepasst werden und die Aufgabe der Transkription und der Alignierung bzw. Segmentierung schneller und unkomplizierter machen.

Bionote: Marcella Palladino is a PhD student at the Università degli Studi di Modena e Reggio Emilia (UNIMORE). Her research focus is on German Linguistics, especially politolinguistics, rhetoric, critical discourse analysis and prosody. During her PhD, she has been Visiting PhD Student at the Institut für Phonetik und Sprachverarbeitung (IPS), LMU in Munich. Her doctoral project deals with the analysis of voice and prosodic characteristics of male and female politicians in Italy and Germany.

Author's address: marcella.palladino@unimore.it

Acknowledgements: I would like to thank PD Dr. Christoph Draxler and the Institut für Phonetik und Sprachverarbeitung (IPS) at the Ludwig-Maximilians-Universität München for their precious support to the current research, and Dr. Daniele Falavigna and his team at the Fondazione Bruno Kessler in Trento for the ASR system used for the Italian speeches.

Literatur

- Brambilla M.M. 2007, *Il discorso politico nei paesi di lingua tedesca: metodi e modelli di analisi linguistica*, Aracne, Roma.
- Burkhardt A., 1996 *Politolinguistik. Versuch einer Ortsbestimmung*, in Diekmannshenke, Hajo/Klein, Josef (Hrsg.), *Sprachstrategien und Dialogblockaden. Linguistische und politikwissenschaftliche Studien zur politischen Kommunikation*, Berlin/New York: de Gruyter, pp. 75-100.
- Draxler C. 2023, *Analysis of transcriptions using Octra – a pilot study*, in Draxler C. (Hrsg.), *Elektronische Sprachsignalverarbeitung 2023*, 105, TUDpress, Dresden, pp. 17-23.
- Fuß. S. and Karbach U. 2019, *Grundlagen der Transkription*, 2. Aufl., Verlag Barbara Budrich, Opladen & Toronto.
- Heilmann C.M. 2003, *Von der Mündlichkeit über die Schriftlichkeit zu veränderter Mündlichkeit. Vom Sinn des Transkribierens*, in Anders L.C., Hirschfeld U. (Hrsg.), *Sprechsprachliche Kommunikation. Probleme, Konflikte, Störungen*, in „Hallesche Schriften zur Sprechwissenschaft und Phonetik“ 12, Peter Lang, Frankfurt/M., pp. 141-151.
- Kirk J.M. and Andersen G. 2016, *Compilation, transcription, markup and annotation of spoken corpora*, in “International Journal of Corpus Linguistics” 21, pp. 291-298. <https://doi.org/10.1075/ijcl.21.3.001int> (15.10.2023).
- Kisler T., Reichel U.D., Schiel F., Draxler C., Jackl B. and Pörner N. 2016, *BAS Speech Science Web Services – an Update on Current Developments*. https://www.researchgate.net/publication/303750698_BAS_Speech_Science_Web_Services_-_an_Update_on_Current_Developments (25.04.2023).
- Kisler T., Reichel U.D. and Schiel F. 2017, *Multilingual processing of speech via web services*, in “Computer Speech & Language” 45, pp. 326-347.
- Niehr T. 2014, *Einführung in die Politolinguistik. Gegenstände und Methoden*, Vandenhoeck & Ruprecht, Göttingen.
- Poerner N. and Winkelmann R. 2018, *Interfacing the BAS Speech Science Web Services and the EMU Speech Database Management System*, in “P&P13 – Proceedings of the Conference on Phonetics and Phonology in German-speaking countries”, Berlin, pp. 141-144.
- Pömp J. and Draxler C. 2017, *OCTRA – a Configurable Browser-Based Editor for Orthographic Transcription*, in “P&P13 – Proceedings of the Conference on Phonetics and Phonology in German-speaking countries”, Berlin, pp. 145-148.
- Schmidt T. 2016, *Good practices in the compilation of FOLK, the Research and Teaching Corpus of Spoken German*, in “International Journal of Corpus Linguistics” 21, pp. 291-298. <https://doi.org/10.1075/ijcl.21.3.001int> (15.10.2023).
- Selting M., Auer P., Barth-Weingarten D., Bergmann J., Bergmann P., Birkner K., Couper-Kuhlen E., Deppermann A., Gilles P., Günthner S., Hartung M., Kern F., Mertzlufft C., Meyer C., Morek M., Oberzaucher F., Peters J., Quasthoff U., Schütte W., Stukenbrock A. and Uhmann S. 2009, *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)*, in „Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion“ 10, pp. 353-402. <http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf> (15.10.2023).

Tools und Software

- Audacity (Version 3.1.3) [Computer software] 1999-2016, Audacity Team. <https://www.audacityteam.org/> (06.02.2024).
- Boersma P., Weenink D. 2022, *Praat: doing phonetics by computer version 6.2.06* [Computer program]. <https://www.praat.org/> (05.12.2023).
- Brugman H., Russel A. 2004, *Annotating Multimedia/ Multi-modal resources with ELAN*, in “Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation”.
- ELAN (Version 6.7) [Computer software] 2023, Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. <https://archive.mpi.nl/tla/elan> (05.12.2023).
- Jochim M., Winkelmann R., Jaensch K., Cassidy S., Harrington J. 2023. *emuR: Main Package of the EMU Speech Database Management System. R package version 2.4.2.* <https://CRAN.R-project.org/package=emuR> (09.02.2024).
- R Core Team 2023. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org> (09.02.2024).
- Schiel F. 1999, *Automatic Phonetic Transcription of Non-Prompted Speech*, in “Proceedings of the ICPhS 1999”, San Francisco, pp. 607-610.
- Schiel F. 2015, *A Statistical Model for Predicting Pronunciation*, in “Proceedings of the ICPhS 2015”, Glasgow, UK, paper 195.
- Schmidt T. 2020, *Exmaralda 1.6* [Computer software]. Universität Hamburg, IDS Mannheim (2011). <https://exmaralda.org/de/ueber-exmaralda/> (05.12.2023).
- Schmidt T., Schütte W. 2010, *FOLKER: An Annotation Tool for Efficient Transcription of Natural, Multi-party Interaction*, in “Proceedings of LREC 2010”. https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/2232/file/Schmidt_Schuette_FOLKER_2010_Paper.pdf (05.12.2023).
- Winkelmann R., Harrington J. and Jänsch K. 2017, *EMU-SDMS: Advanced speech database management and analysis in R*, in “Computer Speech & Language” 45, pp. 392-410.

DIGITALISIERUNG DER POLITISCHEN KOMMUNIKATION

Eine politolinguistische Analyse der Twitter/X- Kommunikation ausgewählter AfD-Politiker

ISABELLA FERRON

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Abstract – This paper investigates the changing nature of German-speaking political communication on social media by analysing a corpus of tweets/posts from the AfD party (*Alternative für Deutschland*). The goal is to find out whether the external communication of some AfD party members on Twitter has changed, if they show any reference to more traditional forms of communication such as public speeches, participation in talk shows, etc., and which linguistic means these party members strategically use to ensure that their messages reach a wider audience of followers. To achieve this, the paper has the following structure: one paragraph (2) describes online political communication, a second (3) focuses on Twitter/X; a third (4) deals with the selection of the corpus and the method used. The last (5) concentrates on the qualitative linguistic analysis of a selection of tweets/posts.

Keywords: political online communication; Twitter/X; AfD-politicians; digitalization.

1. Einleitende Bemerkungen: Fragen und Ziel

Auf Facebook vermischen sich etwa kommerzielle Einträge mit nicht-kommerziellen [...] Der Umgang mit digitalen Identitäten (Avatare); die Ausführung digitaler Kommunikation (die sich alleinig durch ihre Konventionen von anderen Kommunikationsformen unterscheidet); die Erarbeitung einer Netiquette für den Umgang mit anderen virtuellen Identitäten; die Veränderung von Beziehungen, Konsumgütern und Medienrezeption durch soziale Medien; Bewertungsplattformen für unternehmerisches Handeln; Online Dating; der Umgang mit Familien und Freunden auf sozialen Plattformen; Aspekte von Privatheit und Datenschutz; aber auch wie sich Informationen in Netzwerken verbreiten und wer dafür verantwortlich ist oder wie Menschen mit Information Overload, ständiger Erreichbarkeit und dem Streben nach sozialem Status durch Konnektivität umgehen – all diese Phänomene haben direkte praktische Konsequenzen, sei es die Frage, wie viele und welche Daten Menschen bewusst freigeben wollen, können oder müssen; wie sie sich auf sozialen Netzwerken präsentieren und wahrgenommen werden möchten, wie sie Kommunikationsteilnehmer (auch Unternehmen) wahrnehmen bis hin zur Frage, wie sich der Begriff der Freundschaft durch die Überbordung an Statusmeldungen und Informationen verändert hat. (Stiegler 2015, S. 23)

Dieses lange Zitat zu Beginn des Beitrags erläutert, wie stark die Digitalisierung unser Alltagsleben und somit unsere Kommunikationsformen verändert hat. Niemand bezweifelt heutzutage (vgl. Krotz 2001, 2007), dass sie uns vor besondere Herausforderungen stellt, weil die technologischen Innovationen einerseits die Kommunikation erleichtern, andererseits auch die Komplexität der menschlichen Handlungen erhöhen. 2017 definiert Baecker Digitalisierung wie folgt: „Digitalisierung als sozialer und kultureller Prozess heißt, dass sich Maschinen an Kommunikation beteiligen und dass alle anderen Akteure (Menschen, Organisationen, Teams) sich darauf einstellen, dass sie sich beteiligen“ (Baecker 2017, S. 19). Die Digitalisierung senkt die Kosten der Kommunikation und ermöglicht ein Mehr an publizierten Mitteilungen, erschwert jedoch zugleich die Chance gesellschaftlicher Wahrnehmung und gelungener Kommunikation (vgl. Krotz 2007, S. 209). Durch die algorithmische Vermittlung können Akteure ihre Botschaften in höheren Auflösungen an spezifische Zielgruppen richten und sich mit ihnen verbinden (vgl. Beißwenger, Storrer 2012, S. 92-124).

Die Digitalisierung verändert nicht nur die Alltagskommunikation: Als Bestandteil und Treiber eines grundlegenden strukturellen Wandels (vgl. Vowe 2020, S. 4) beeinflusst sie auch die politische Kommunikation. Letztere wird in diesem Zusammenhang nicht nur als die Kommunikation von Politikern¹ mit ihren Wählern, also als externer politischer Dialog verstanden, sondern bezieht auch den journalistischen Diskurs um Politik mit ein, der sich durch die neuen Medien im Netz verbreitet. Durch die Massenmedien entsteht eine besondere Form von politischer Kommunikation, die von ihrer Simultaneität profitiert und ein neues Verhältnis zu den potenziellen Wählern bzw. zu der Bevölkerung entwickelt. Diese Kommunikation ist in ihrem Ausdruck zugleich effizienter, pluraler, individueller, globaler, dynamischer und hybrider und ermöglicht eine höhere Teilnahme an öffentlichen Debatten.

Seit den 2000er-Jahren sind drei Tendenzen im Verständnis von Digitalisierung und ihrem Einfluss auf die politische Kommunikation erkennbar: Die erste sieht ihren fundamentalen Wandel in Form eines Bruchs von bisherigen Strukturen (Revolutionsthese); die zweite behauptet, dass nicht etablierte oder periphere politische Akteure von der Digitalisierung profitieren würden, um bekannter zu werden; die dritte argumentiert, dass es zu Veränderungen komme, aber nach einer Weile die politischen Kräfte „normale“ Verhältnisse wieder herstellen würden. Zu beachten ist aber auch, dass Online-Medien nicht nur zur Vereinfachung und Verbesserung des gesellschaftlichen Diskurses beitragen, sondern in der Realität bestehende gesellschaftliche Spaltungen fortschreiben (vgl. Emmer *et al.* 2011), da Machtasymmetrien, Ausbeutung digitaler Arbeit, Überwachung durch

¹ Aus Gründen einer besseren Lesbarkeit wird das generische Maskulinum gebraucht. Die personenbezogenen Bezeichnungen gelten als geschlechtsneutral.

Unternehmen und Staat usw. (vgl. Fuchs 2021, S. 9-26) auch online stattfinden. Die in den letzten Jahren geführte Debatte um die Metaphern der Filterblasen und Echokammer, mit denen Aspekte des Phänomens der Mikroöffentlichkeiten erklärt werden, zeigt, dass diese nur eine Facette der Komplexität politischer Meinungsbildungsprozesse durch Online-Medien beschreiben (vgl. Stark *et al.* 2021, S. 303-321).

In dieser Hinsicht zielt der vorliegende Beitrag darauf ab, die deutschsprachige politische Kommunikation auf der Microblogging-Plattform Twitter (seit August 2023 X genannt) sprachwissenschaftlich durch die Analyse eines Korpus aus Tweets² von einigen AfD-Parteimitgliedern zu untersuchen. Allgemein wird Twitter/X von Politikern verwendet, um ihre persönlichen Einstellungen zu verschiedensten Ereignissen zu äußern. Neben Pressekonferenzen und Pressemitteilungen ist Twitter/X ein Kommunikationskanal, der Politikern zur Verfügung steht, um z. B. ihre eigene Agenda zu gestalten und die öffentliche Meinung zu beeinflussen. Ziel des Beitrags ist es zu sehen, ob sich die externe Kommunikation von AfD-Parteimitgliedern auf Twitter verändert hat, ob diese Kommunikationsart Berührungspunkte mit traditionelleren Kommunikationsformen (öffentliche Reden, Teilnahme an Talkshows usw.) zeigt, und welcher sprachlichen Mittel sich diese Parteimitglieder strategisch bedienen, damit ihre Mitteilungen ein breiteres *Followers*publikum erreichen. Twitter/X eignet sich – so die zugrunde liegende These – durch die konzeptionell mündliche Form und die begrenzte Zeichenzahl für populistische Äußerungen, wie sie für AfD-Parteimitglieder typisch sind (vgl. Girnth 2013, S. 119-120).

Die Arbeit ist in folgende Abschnitte unterteilt: Ein Abschnitt (2) definiert die politische Online-Kommunikation, ein zweiter (3) befasst sich mit Twitter/X; ein dritter (4) erläutert die Auswahl des Korpus und die verwendete Methode. Der letzte (5) fokussiert sich auf die qualitative sprachliche Analyse einer Auswahl von Tweets/Posts.

2. Politische Online-Kommunikation

Die politische Kommunikation gehört zu den Grundelementen der Demokratie, wodurch politische Akteure einerseits als Sender agieren, um die Bevölkerung zu informieren; andererseits gelten sie auch als Rezipienten des Feedbacks der Bevölkerung, das sie zu beachten haben, um diese Bevölkerung vertreten zu können. Nach Sarcinelli (2009, S. 85-86) ist Kommunikation allgemein als eine Form von Legitimation der Politik zu sehen; laut Eilders

² Mit der Umbenennung der Plattform von Twitter in X wird nicht mehr von Tweets, sondern von Posts gesprochen. In diesem Beitrag werden beide Begriffe verwendet, allerdings mit einer Anmerkung: Bis zum Abschnitt 3, der eine kurze Einführung in die Plattform gibt, wird die Bezeichnung Tweet/Post verwendet. In den folgenden Abschnitten wird Post benutzt.

(2006, S. 283) umfasse sie alle Prozesse der Informationsvermittlung, die sie zum Gegenstand hat. 1990 stellt Wolton die politische Kommunikation in eine räumliche Dimension: „the space in which contradictory discourse is exchanged between three actors with the legitimate right to express themselves in public on politics, namely politicians, journalists and public opinion by means of opinion polls“ (Wolton 1990, S. 12). Wolton unterscheidet ferner zwischen einer politischen Debatte, in der es um Konsensgewinn geht (ebd.), und der öffentlichen Arena, die alle Formen politischer Äußerungen versammelt. Zu dieser öffentlichen Arena gehört die Kommunikation auf Online-Plattformen.

In ihrer wechselseitigen Beziehung zur Bevölkerung und zu den Medien wird politische Kommunikation von Jarren und Donges (2002, S. 20) als ein gesellschaftliches Totalphänomen verstanden, das sich nur schwer in einzelne Teilaspekte aufteilen lässt. Sie ist durch die Massenmedien mit einem Konzept von Öffentlichkeit verbunden, das von dem von Habermas 1962 geprägten ausgeht, jedoch (vgl. Bruns, Highfield 2016, S. 53-56) einem stetigen Wandel unterliegt und eine Komplexität aufweist, die nicht von einer Elite reguliert wird. Vielmehr zählt sie zur Fragmentierung eines einzigen Öffentlichkeitsraums oder besser einer „networked public sphere“ (Benkler 2006, S. 212). Im Jahr 2024 konstruiert sich zwar Öffentlichkeit nicht mehr über das medienöffentliche Handeln etablierter politischer und gesellschaftlicher Akteure, sondern wird durch eine enthierarchisierte und direktere Kommunikation verstärkt. Voraussetzung dafür ist die Digitalisierung, deren Bedeutungsgewinn in Medien und Politik immer größer wird. Die politische Online-Kommunikation kann wie im Folgenden charakterisiert werden:

1. Wie das politische System selbst, das von einem ständigen Veränderungsprozess sowohl auf nationaler als auch internationaler Ebene abhängt, zeigt die politische Online-Kommunikation eine große Anpassungsfähigkeit an die Flexibilität der Kommunikations- und Mediensysteme (vgl. Jarren, Sarcinelli 1998, S. 13-20).
2. Sie steht vor der Herausforderung, einen Kompromiss zwischen Darstellungspolitik und Entscheidungspolitik zu finden (vgl. Korte, Hirscher 2001, S. 11-12), d. h. zwischen zwei sich unterscheidenden politischen Kommunikationswelten mit eigener Funktionslogik. Während Entscheidungspolitik auf strategische Kommunikation angewiesen ist, konzentriert sich die Darstellungspolitik auf die politische Momentaufnahme; sie bildet und artikuliert sich durch die Medien in einer für alle zugänglichen Sphäre. Die Interaktion durch die Medien gestattet ferner die Integration der unterschiedlichen Bevölkerungsgruppen (integratives Vermögen von Medienkommunikation; vgl. Deppermann 2018, S. 51-87).

3. Sie zeigt und erlaubt Interaktion: Während eine einseitige Kommunikation in traditionellen Medien wie Fernsehen stattfindet, bietet das Internet die Möglichkeit eines Rückkanals zum Sender. Somit kann die Bevölkerung in Form von Diskussionen und Kommentaren am politischen Prozess teilhaben (vgl. Busemann, Gscheidle 2009, S. 356).
4. Sie ermöglicht die Personalisierung vom Image des jeweiligen Politikers in Videoblogs, Livestreams usw.
5. Sie gestaltet die mediale Agenda eines jeweiligen Politikers, der somit ein sehr breites Publikum erreichen kann (vgl. Feezell 2018, S. 482-494; Gilardi *et al.* 2022, S. 39-60).
6. Politische Online-Kommunikation funktioniert gut, wenn die rechtlichen Vorgaben eingehalten werden. Dies verringert auch die Gefahr der Verbreitung von Fake News.
7. Sie gewinnt eine Durchsetzungsfähigkeit im Aufmerksamkeitswettbewerb (vgl. Sarcinelli 2009, S. 196), d. h. im Versuch, die jeweilige parteiliche Außendarstellung zu medialisieren und somit zu legitimieren.

Von einem sprachwissenschaftlichen Standpunkt aus betrachtet ist vorwiegend die Pragmatik des sprachlichen Handelns von Interesse. Eine sprachwissenschaftliche Beschäftigung mit dieser Art von Sprachgebrauch existiert seit den frühen 1990er-Jahren (vgl. Haase *et al.* 1997, S. 51-85; Holly 2011, S. 144-163; Runkel *et al.* 1998, S. 97-109), wird in den 2000er-Jahren weiter vertieft und als *Internet Linguistics* betitelt (Crystal 2005): Diese Definition erweist sich jedoch als problematisch und unscharf, weil *Internet Linguistics* unterschiedliche Kommunikationsformen (E-Mails, Chats, Blogs, Tweets/Posts usw.) untersucht. Demzufolge muss man jeweils die Kommunikationsform präzisieren und den methodischen Ansatz anpassen. In dieser Hinsicht sind Tweets/Posts kurze Texte, deren Haupteigenschaft in ihrer Kürze liegt, die eine Mitteilung greifbar macht. Sie sind nicht nur hinsichtlich der Pragmatik des sprachlichen Handelns bedeutend, sondern auch auf lexikalischer und semantischer Ebene. Ferner sind sie Teil einer Kommunikationsform, die als Vermittlung zu verstehen ist. Nach Luhmann ist Kommunikation als eine Synthese aus drei Selektionsprozessen zu verstehen: Information, Mitteilung und Verstehen. Eine Information ist „nach heute geläufigem Verständnis eine Selektion aus einem (bekannten oder unbekanntem) Repertoire an Möglichkeiten“ dessen, was kommuniziert werden kann und was nicht (Luhmann 1984, S. 95). Eine Mitteilung ist die absichtliche/unabsichtliche Wahl eines Verhaltens, das diese Information mitteilt (ebd.). Verstehen ist schließlich die „Erwartung einer Annahmeselektion“ (ebd., S. 196). Dabei können zwei Formen dieses Verhaltens unterschieden werden: Übermittlung und Vermittlung. Jede Mitteilung wird in Form einer sprachlichen Äußerung oder eines Textes übermittelt; darauf folgt eine Vermittlung von Bedeutungen, eine soziale

Konstruktion von Sinn (vgl. Beck 2020, S. 33). Vermittlung ist ein komplexer Prozess, der vielfältige Elemente einbezieht: die Differenz von Information und Mitteilung, die Kommunikationsbeteiligten, das verwendete Zeichensystem oder die Handlungssituation sowie damit verbundene Kompetenzen.

3. Twitter/X

Twitter/X ist eine Microblogging-Plattform, auf der registrierte Nutzer (*Follower*) Kurznachrichten (*Tweets/Posts*) veröffentlichen (vgl. Storrer 2009, S. 2211-2226; O'Reilly, Milstein 2009; Murthy 2013). Ein Tweet/Post darf maximal 280 Zeichen lang sein: Tweets/Posts erscheinen auf der Startseite in umgekehrter chronologischer Reihenfolge (*Timeline*), d. h., die neuesten Mitteilungen stehen ganz oben. Wird eine neue Nachricht gepostet, rücken alle anderen entsprechend nach unten. Ein Tweet/Post ist eine multimodale Mitteilung, die durch Fotos und/oder Videos ergänzt oder durch Links erweitert werden kann. Daraus eröffnet Twitter/X eine Vielzahl von Nutzungsformen, die vom einfachen Mitteilungstweet bis zum komplexen narrativen Tweet/Post reichen: Tweets/Posts können z. B. eine Art Tagebucheintrag mit persönlichen Kommentaren, einer Sammlung von Fotos, Sprüchen, Witzen sein; sie können offiziell Institutionen oder soziale Gruppen vertreten, Live-Reportagen, Zeitungen und Sendungen repräsentieren usw.

Als Twitter-Nutzer kann man einerseits anderen Nutzern folgen (A folgt B, aber B folgt nicht notwendigerweise A), deren Beiträge auf der eigenen Startseite sehen, darauf antworten, sie kommentieren oder weiterleiten (*retweet/repost*). Andererseits hat man als Autor auch Follower, die die eigenen Tweets/Posts lesen (vgl. Moraldo 2009, S. 245-281; Moraldo 2012a, S. 77-85, Moraldo 2012b, S. 2-13). Wie andere Online-Plattformen kann Twitter/X als „ein sozialitätsstiftendes Medium“ definiert werden, wobei man dadurch soziale Bedürfnisse offenbart (vgl. Java *et al.* 2007, S. 56-65). Als vielseitiges Medium, das vorwiegend eine informative Funktion erfüllt, ermöglicht es ein schnelles Update verschiedener öffentlichkeitsrelevanter Situationen oder auch aktueller politischer Entwicklungen (z. B. Ukraine-Krieg, Terroranschläge der Hamas seit dem 7. Oktober 2023).

Auf Twitter/X vollzieht sich die Kommunikation durch sogenannte Operatoren, die als semiotisch bezeichnet werden (vgl. Thimm *et al.* 2011, S. 265-285) und eine Art kollaborativer Aktivität erlauben (vgl. Stephens 2010, S. 204). Sie vermitteln bestimmte Inhaltsbeschreibungen und wirken konnotativ:

- @: Das At-Zeichen hat die kommunikative Funktion, etwas bzw. jemanden zu adressieren, zu erwähnen. Es gehört zu einer diskursiven Strategie der Gesprächspartnerauswahl (vgl. Herring, Honeycutt 2009, S. 12).

- # (Hashtag): Es strukturiert die Nachricht. Es gilt als Schlüsselwort, weil es ein bestimmtes Thema hervorhebt und in der Regel von einem Substantiv oder einer Wortgruppe begleitet wird (z. B. #NieWieder). Es hat eine kommunikative Funktion, weil Einträge dadurch mit Begriffen versehen werden, die die Botschaft des Tweets/Posts in einen neuen Kontext stellen. Daher trägt seine Verwendung zur „Konturierung, Zuspitzung und Emotionalisierung der jeweiligen Diskurse bei und setzt dabei auf individuelle, parteiliche, ideologische oder politische Unterscheidungssignale“ (Niehr 2007, S. 498; vgl. auch Gilbert 2001, S. 239-249).
- <http://>: Der Link hilft bei der Referenz eines Themas und seiner multimodalen Weiterführung.
- ↻ (Retweet/Repost): Diese Funktion dient der Redistribution von Tweets/Posts und zeigt das Potenzial von Twitter/X als soziales Kommunikationsgefüge, in dem Nutzer über Dritte miteinander verbunden sind.

Da Twitter/X sowohl für private als auch für institutionelle Follower Optionen der Nutzung im Sinne von Kommunikation, Interaktion, Diffusion und Partizipation bietet, hat es politische Implikationen: Es kann Einfluss auf partizipative Prozesse und Kommunikationsstrukturen haben. Die Folgefunktion von Politiker-Twitterprofilen seitens der Bürgerschaft erlaubt es sozusagen, diese Politiker in ihren Aktivitäten zu kontrollieren. Die Funktionen von Retweet/Repost, @-Kommunikation, Hashtags und Links lassen vielfältige Bezugnehmehandlungen ausführen. Vor allem ermöglicht die Verwendung von Hashtags eine Partizipation beim Online-Publizieren.

Der Sprachgebrauch auf Twitter/X weist sowohl Merkmale der Mündlichkeit als auch der Schriftlichkeit auf; es geht um eine zugleich sprechsprachlich und medial bedingte Schriftlichkeit, d. h. – mit der Definition von Koch und Oesterreicher – um eine Sprache der Nähe und zugleich der Distanz (Koch, Oesterreicher 1985, S. 15-43; Koch, Oesterreicher 2008, S. 199-215; Moraldo 2012c, S. 183-208; Thaler 2003). Tweets/Posts gehören zur Sprache der Distanz, weil sie schriftlich vermittelt und somit immer abrufbar sind; andererseits gleicht ihr Stil demjenigen der Mündlichkeit, d. h. der Sprache der Nähe, wenn man z. B. an emotionale Ausrufe, kurze Zwischenfragen oder Sätze, Wiederholungen und Redundanzen oder an fließende Übergänge zwischen rein standardsprachlichen und dialektalen Strukturen denkt.

4. Korpus und Methode

4.1. Korpus

Das untersuchte Korpus (Korpus_postX_AfD) besteht aus 557 Posts (insgesamt 162 800 Wörter) der folgenden Twitterprofile von AfD-Politikern im Zeitraum Mai 2023 bis Januar 2024: 1. das offizielle AfD-Parteiprofil, 2. das Profil von Alice Weidel, 3. das Profil von Beatrix von Storch sowie die Profile von 4. Tino Chrupalla und 5. Björn Höcke, die als die repräsentativsten Politiker der Partei gelten. Am wortreichsten und am meisten präsent im Posten ist Beatrix von Storch, die täglich mehrmals Posts absetzt.

Die Posts wurden manuell über die *Timeline* der Profile gesammelt: Sie betragen regelmäßig ca. 280 Zeichen; eine Ausnahme ist Höcke, der neben den regulären Posts auch längere Mitteilungen postet, wozu man ein X Blue-Abonnement³ benötigt. Die Verteilung der Posts auf die Profile ist wie folgt:

Profile	Postsanzahl (insgesamt 557)
AfD	90
Beatrix von Storch	212
Alice Weidel	120
Björn Höcke	80
Tino Chrupalla	55

Tabelle 1
Metadaten von Korpus_postX_AfD.

Die Posts lassen sich in vier Kategorien einteilen, die sich nicht immer klar voneinander unterscheiden, sondern oft auf den Profilen der Politiker koexistieren:

- (a) Der betreffende Politiker erzählt entweder in deskriptiver oder polemischer Form, was an einem typischen Tag passiert ist; bisweilen wird von der Teilnahme einer öffentlichen Veranstaltung berichtet. Diese Posts werden in der Regel von Fotos oder Videos begleitet:

³ Ein X Blue-Abonnement hat für Abonnenten weitere Funktionen, wie z. B. die Möglichkeit, längere Posts zu posten.



Abbildung 1
Post von Alice Weidel, 19.09.2023.

(b) Der jeweilige Politiker unterstützt eine Partei bzw. einen Politiker bei Regionalwahlen;



Abbildung 2
Post von Tino Chrupalla, 02.07.2023.

(c) Der jeweilige Politiker nimmt zu einem bestimmten historischen oder aktuellen Ereignis Stellung; es kann sich auch um Repost von Posts von Parteikollegen, Journalisten usw. handeln, die kommentiert werden:



Abbildung 3
Post von Beatrix von Storch, 27.10.2023.

(d) Der jeweilige Politiker tritt gegen einen politischen Gegner auf, diskreditiert oder verspottet ihn:



Abbildung 4
Post von Beatrix von Storch, 08.11.2023.

4.2. Methode

Was die Methode betrifft, liegt der Schwerpunkt vorwiegend auf der Theorie der Sprachhandlungen (vgl. Austin 1962; Searle 1981; Krämer 2001), die in der Analyse der politischen Online-Kommunikation von Bedeutung sind, weil sie ihren performativen und pragmatischen Charakter hervorheben.

Den Äußerungen politischer Akteure auf Twitter/X liegt ein Kooperationsprinzip zwischen Redner und Leser zugrunde, das in der öffentlichen Kommunikation die Verbreitung und Rezeption der Botschaft ermöglicht (vgl. Niehr 2021, S. 75-85; Marcinkowski 2014, S. 7-17). Die Theorie der Sprachhandlungen erlaubt es schließlich, nicht nur die sprachlichen Merkmale der Posts zu untersuchen, sondern sie auch mit Bezug auf die politische Semantik präziser zuzuordnen und zu verstehen. Bei den Posts der AfD-Politiker handelt es sich um assertive Sprechakte, weil sie einen Blick auf Deutschland und die Welt werfen, sie beschreiben und eine wahre bzw. falsche Aussage über sie ausdrücken. Sie sind auch expressiv (vgl. Searle 1981, S. 15), weil sie durch ein großes Emotionspotential charakterisiert sind. Demzufolge sind sie sachverhaltszentrierte Sprechakte, bei denen ein Sachverhalt mit einer Sprecherbewertung verknüpft wird. Zugleich handelt es sich aber auch um gegnerbezogene Sprechakte, in denen eine emotionale Haltung des jeweiligen Followers gegenüber einem politischen Gegner ausgedrückt wird.

Die Untersuchung bezieht auch die anderen Sprachebenen (Lexik, Syntax usw.) ein, vor allem die lexikalische Ebene: Auf Twitter/X werden oft Akronyme und Abkürzungen verwendet, die die Geschwindigkeit der Beiträge beschleunigen. Auch Ellipsen und Majuskeln sind häufig zu verzeichnen, die zur Erhöhung der zu imaginierenden Lautstärke oder zur Emphase beitragen sollen (vgl. Haase *et al.* 1998, S. 68).

5. Qualitative Analyse der Posts

Durch die sprachliche Analyse der Posts soll gezeigt werden, welche kommunikativen Strategien diese Politiker verwenden, um ihre Ziele zu erreichen. Es ist vorab zu bemerken, dass die ausgewählten Posts in verschärfter und konzentrierter Form den mündlichen Diskursstil zeigen, den diese Parteivertreter in ihren öffentlichen Reden auf unterschiedliche Art und Weise pflegen. Oft ist der Stil salopp und informell, der gesprochenen Sprache nah, dann wieder formell, wenn die Rede von der Regierung und z. B. von parlamentarischen Debatten ist. Sowohl das Parteiprofil als auch die Profile der vier Politiker bedienen sich der Abkürzungen, Ellipsen und Akronyme. Sie bilden kreative Schlagwörter oder Schlagwortkombinationen und gebrauchen Okkasionalismen oder einzelne Lexeme in Form von Hashtags, wenn sie z. B. einen politischen Gegner diskreditieren wollen.

Die Posts setzen von den Followern ein gewisses Wissen dessen voraus, was aktuell in Deutschland und weltweit passiert. Das ist vor allem in den Hashtags zu beobachten, die sich auf besondere Ereignisse beziehen (Bauernproteste, Haushalt, Ukraine-Krieg usw.). Unabhängig von dem im Zentrum des jeweiligen Posts stehenden Themas (Innen- und Außenpolitik, deutsche Gesellschaft, Migration, Islam usw.) kann man folgende sprachliche und bildliche Aspekte bemerken:

- Die Politiker, vor allem Beatrix von Storch, haben gute Kenntnisse der Twittersprache: In (1) bezieht sich von Storch auf das spezifische Problem der Sanierung eines Seniorenheims in Berlin und verbindet es mit der Flüchtlingsproblematik, um somit kritische Aspekte in der politischen Agenda der Regierung hervorzuheben, z. B. die Vernachlässigung der eigenen Bevölkerung zugunsten der Flüchtlinge (*Die Bürger kommen an letzter Stelle*). Das wird auch durch den Gebrauch von den sich semantisch widersprechenden Verben *ausziehen* und *einziehen* betont. Der Post zeichnet sich durch einen sarkastischen letzten Satz aus, in dem er durch die epistemische Verwendung des Modalverbs *müssen* und des zusammengesetzten Wortes *Cum-Ex-Scholz* auf einen Steuerhinterziehungsskandal aufmerksam macht. Dieser liegt mehr als fünf Jahre zurück, als Scholz noch Bürgermeister der Stadt Hamburg war, und in den auch er verwickelt war, wie es scheint. Der Terminus *Cum-Ex* ist ein Fachterminus, der ein System des Handels mit Aktien mit (*cum*) oder ohne (*ex*) Dividendenanspruch beschreibt, nachdem ein Unternehmen den Zinssatz zwar angekündigt, aber nicht ausgezahlt hat. Durch diesen Beinamen diskreditiert von Storch den Bundeskanzler zweifach: Einerseits, weil sie dem Bundeskanzler rechtswidrige Handlungen unterstellt; andererseits soll das Wort in Verbindung mit dem Hashtag *#Respekt* die Gleichgültigkeit der Regierung den deutschen Bürgern gegenüber beweisen:

(1)

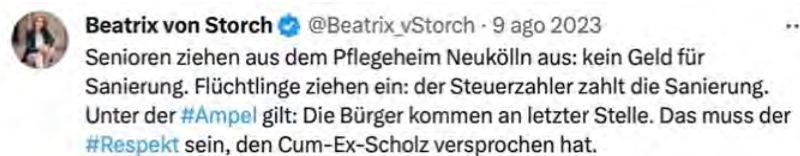


Abbildung 5
Post von Beatrix von Storch, 09.08.2023

- Oft verwenden die untersuchten Politiker Spitznamen, die sie erfunden haben, um den Gegner zu diskreditieren. Solch pejorative und abwertende Spitznamen (vgl. Finkbeiner *et al.* 2016, S. 1-18) werden vor allem für den Bundeskanzler und die Regierung verwendet. In (2) und (3) äußern Beatrix von Storch und Tino Chrupalla ihre Meinungen über den Kanzler:

(2)



Abbildung 6
Post von Beatrix von Storch, 10.01.2024

Von Storch benennt ihn mit dem Beinamen *Respektkanzler* und verbindet ihn oxymorisch mit dem Verb *versprechen*, um aufzuzeigen, dass er beim Thema „Wohnungen“ nichts von dem umgesetzt hat, was er während der Wahlkampagne versprochen hatte. Dass es nicht nur um den Kanzler geht, sondern dass er stets mit seinen verfehlten Handlungen für die von der Regierung in Gang gesetzten Maßnahmen steht, wird durch den Gebrauch der Redewendung *gegen die Wand fahren* im letzten Satz hervorgehoben.

In (3) vergleicht ihn Chrupalla mit einem Piraten, um in der Generaldebatte Scholz' Stellungnahme zu beschreiben, die nach Chrupalla negative Wirkungen für Deutschland haben wird. Solch eine Meinung wird auch durch den Gebrauch des Verbs *rauben* ausgedrückt, wodurch Chrupalla den widerrechtlichen Charakter dieser Handlung betont (vgl. Technau 2018, S. 3, 69-70, 337-340).

(3)



Abbildung 7
Post von Tino Chrupalla, 06.09.2023

- Hashtags werden in der Regel in Zusammenhang mit Schlagwörtern (#Kartellpartei) oder ad hoc gebildeten Ausdrücken (#DeshalbAfD#nurnochAfD) eingesetzt (4), sowie auch mit Symbolen wie der deutschen Fahne oder Emoticons (5):

(4)



Abbildung 8
Post von Beatrix von Storch, 27.11.2023

In (4) verwendet von Storch im Hashtag eine politikwissenschaftliche Bezeichnung (*Kartellparteien*), mit der Parteiforscher wie Richard S. Katz und Peter Mair (Mair, Katz 1995, S. 5-28) in den 1990er-Jahren den Wandel der Parteien in Westeuropa beschrieben haben. Von Storch bezieht sich auf eine bestimmte Form von Politik, die aber nicht mehr in der Lage ist, zum Nutzen der Bevölkerung zu regieren. Das wird auch durch die konzessive Konjunktion *obwohl* betont.

In (5) hilft die Verbindung von Hashtag, Symbol sowie der deutschen Fahne, die Wichtigkeit der Mitteilung zu verstehen. Die schwierige Lage Deutschlands mit Bezug auf die Kernenergie wird auch durch die Metapher des „Geisterfahrers“ (umgangssprachlich für „Falschfahrer“) betont. Zu bemerken sei ferner, dass Symbole, Emoticons usw. in Posts verwendet werden, die aus kurzen Nominalsätzen und Ausrufesätzen bestehen und einen umgangssprachlichen Ton haben.

(5)



Abbildung 9
 Post von Beatrix von Storch, 03.12.2023

- Großgeschriebener Wörter bedient sich unter den fünf untersuchten Profilen nur Beatrix von Storch, um einen besonderen Aspekt in ihren Posts hervorzuheben:

(6)



Abbildung 10
 Post von Beatrix von Storch, 23.08.2023

- Im Vergleich zu Beatrix von Storch und Tino Chrupalla, sowie zum Profil der Partei nutzen Björn Höcke und Alice Weidel eine weniger mündlich geprägte Sprache. Ihre Posts ähneln vielmehr der Schriftsprache sowohl wegen der Länge als auch des Registers. Sie sind voll von intertextuellen Bezügen zur Geschichte Deutschlands, die somit für nicht deutsche Follower schwer zu verstehen sind. In (7) nutzt Höcke den Jahrestag von Luthers Thesenanschlag in Wittenberg als Ausgangspunkt für eine Anspielung auf die aktuelle Situation Deutschlands und als implizite Äußerung seiner Polemik gegen Fremde in Deutschland. Der *kleine* Mönch Luther – im Sinne von damals noch nicht einflussreich – wird im ersten Satz zu einem Helden Deutschlands erhöht, der im Namen der Freiheit und gegen alle Formen der Gewaltherrschaft (*Anmaßung und Degeneration*) gehandelt habe:

(7)



Abbildung 11
Post von Björn Höcke, 31.10.2023

Wie Luther in (7) wird noch eine bedeutende Persönlichkeit der deutschen Kulturgeschichte, Wilhelm Busch, in (8) von Höcke gebraucht, um den Erfolg der AfD bei der Oberbürgermeisterwahl in Nordhausen zu verdeutlichen. Wilhelm Busch, Autor der Bildergeschichte *Max und Moritz* (1865), wird bei Höcke als *Meister der dt. Sprache* beschrieben und in Verbindung mit dem Kompositum *Bildungsrettungspartei* gebracht, um – möglicherweise als Erwiderung auf die an der Partei geübte Kritik, sie wolle keine inklusive Schule – zu betonen, dass die AfD die einzige Partei sei, die sich um die Bildung der deutschen Bevölkerung bemühe:

(8)

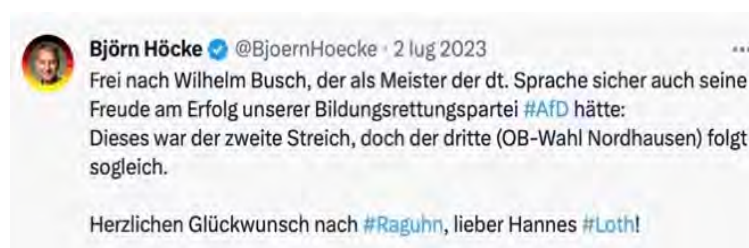


Abbildung 12
Post von Björn Höcke, 02.07.2023

In (9) verwendet Weidel den Konjunktiv I, der vor allem in der Schriftsprache, in Zeitungsartikeln und Nachrichtenberichten gebraucht wird, um Fakten neutral zu berichten. Somit distanziert sie sich von den Leuten, die eine bestimmte Meinung über Robert Habeck haben, und kann ihre Meinung äußern. Dazu diskreditiert sie Habeck in seiner Rolle als Politiker, weil sie ihn als *Mann* bezeichnet und Auszüge aus seinen Reden kommentiert.

(9)



Abbildung 13
Post von Alice Weidel, 28.10.2023

- Oft enthalten die Posts Gegensätze, Oxymora oder onomatopoetische Wörter mit dem Ziel, den Sinn der Mitteilung zu bekräftigen. In (10) schildert Chrupalla das seiner Meinung nach gleichgültige Verhalten der Grünen der Umwelt gegenüber: Er stellt die deutsche Landschaft, die er positiv beschreibt (*Ferienparadies*, *romantisches Idyll*), dem Import von Fracking-Gas gegenüber, das durch das Adjektiv *dreckig* negativ konnotiert ist:

(10)

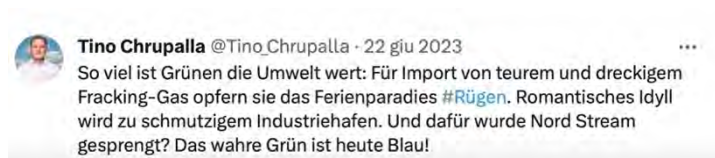


Abbildung 14
Post von Tino Chrupalla, 22.06.2023

In (11) verwendet von Storch das onomatopoetische Wort *Doppelwumms*: „Wumms“⁴ (auch „Bums“) beschreibt einen sehr lauten Knall, der die Menschen erschreckt. Im übertragenen Sinn beschreibt das Wort, das hier in seiner Bedeutung durch den Modifikator *Doppel-* als Übertreibung bekräftigt wird, ein einschneidendes Ereignis, das aufschreckt:

⁴ Vgl. *Wumms* in Digitales Wörterbuch der deutschen Sprache (DWDS), <https://www.dwds.de/wb/Wumms>, (letzter Zugriff: 07.07.2024).

(11)

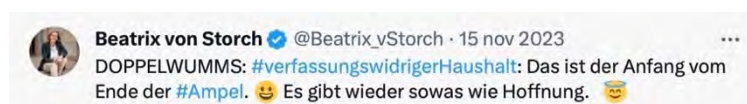


Abbildung 15
Post von Beatrix von Storch, 15.11.2023

- Posts enthalten oft Klimaxe, um die Problematik einer bestimmten Situation zu verdeutlichen: In (12) beschreibt Höcke eine Reihe von Fakten bzw. Ereignissen, die sich erst kürzlich ereignet haben. Da es sich um sozial gefährliche Sachverhalte (*Ausschreitungen, Messerstecherei*) handelt, werden sie als Folge der verfehlten Sozialpolitik der früheren Regierungen gewertet. Um das zu bekräftigen, verwendet Höcke veraltete Wörter wie *Randale* und den Aufruf *Willkommen in der schönen neuen Welt der Altparteien*, indem er mit den Wörtern *neu* und *Altpartei* spielt. Die Bezeichnung *Altpartei* beschreibt eine Partei mit einer langen parlamentarischen Tradition, kann im politischen Sprachgebrauch jedoch eine pejorative Bedeutung haben, wenn der Begriff wie in diesem Post als Schlagwort verwendet wird. Das Wort findet man häufig bei AfD-Politikern, vor allem bei Höcke, der es als abwertenden Sammelbegriff zur Bezeichnung anderer Parteien gebraucht. Die abwertende Bedeutung dieses Begriffs wird vom sich auf die Welt beziehenden Adjektiv *neu* unterstrichen, das mit gewissem Sarkasmus darauf hinweist, dass die traditionelle Politik keine Probleme beseitigen kann. Ferner spielt er auch auf Aldous Huxleys Roman *Brave New World* (1931) an:

(12)



Abbildung 16
Post von Björn Höcke, 17.09.2023

- Häufig zeigt sich die Verwendung von Redewendungen, Metaphern oder Vergleichen (13), um die Probleme stärker zu betonen, vor allem mit Bezug auf die Migrationspolitik, die der deutschen Gesellschaft unterschiedliche Probleme verursacht hat. In (13) bezieht sich von Storch auf das Leben der deutschen Juden, das ihrer Meinung nach nicht von rechten Bewegungen erschwert wird, sondern von Migranten muslimischer Religion. Um die Aufmerksamkeit auf ihren Diskurs zu lenken, veröffentlicht sie mit dem

Post das Video ihrer Rede im Bundestag, in dem der Ausdruck *Der Elefant im Raum* weiß auf grün geschrieben ist. Im Post verwendet sie ferner das Hashtag *#muslimisch*. „Der Elefant im Raum“ (oder auch „Der Elefant im Zimmer“) ist eine Metapher, die ursprünglich im russischen, heute im englischsprachigen Raum verbreitet ist. Damit wird ein Problem bezeichnet, das für eine Gruppe von Menschen klar erkennbar ist, jedoch nicht thematisiert wird. In diesem Fall stellt der Elefant die Muslime dar, die in Deutschland wohnen und sowohl für die deutsche Bevölkerung als auch für den jüdischen Teil derselben eine Gefahr andeuten sollen:

(13)



Abbildung 17
Post von Beatrix von Storch, 19.10.2023

In (14) nutzt Weidel den Vergleich *wie ein Fisch im Wasser*, um das Verhalten von Gewalttätern zu beschreiben, das als Folge einer unkontrollierten Einwanderung zu verstehen ist:

(14)

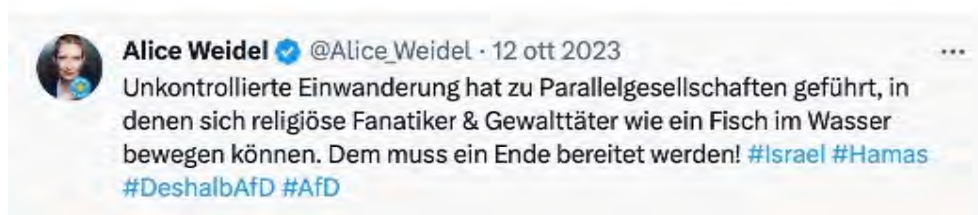


Abbildung 18
Post von Alice Weidel, 12.10.2023

Die folgende Tabelle fasst die sprachlichen Strategien der AfD-Mitglieder im untersuchten Post-Korpus zusammen und gibt die Anzahl der Posts an, in denen sie verwendet werden. Dadurch wird ersichtlich, welcher Politiker sie am häufigsten und regelmäßig verwendet, sodass sie zu einem festen

Bestandteil des persönlichen Stils werden.

Verwendete sprachliche Strategien	Twitter/X Profile				
	AfD-Profil	Beatrix von Storch	Alice Weidel	Tino Chrupalla	Björn Höcke
Diskreditierung der politischen Gegner durch Spitznamen	40 Posts	112 Posts	76 Posts	30 Posts	55 Posts
Sarkasmus		77 Posts	40 Posts	13 Posts	45 Posts
Hashtag mit Schlüsselwörtern, Symbolen, Emoticons	90 Posts	212 Posts	100 Posts	45 Posts	75 Posts
Intertextuelle Bezüge	45 Posts		83 Posts		75 Posts
Großschreibung		105 Posts			
Mündliches Register	35 Posts	100 Posts		20 Posts	
Schriftliches Register	55 Posts			10 Posts	78 Posts
Gegensätze, Oxymora, onomatopoetische Wörter	15 Posts	130 Posts	70 Posts	25 Posts	53 Posts
Rhetorische Figuren	22 Posts	170 Posts	95 Posts	17 Posts	45 Posts

Tabelle 2
Sprachliche Strategien der AfD-Politiker in Posts.

Die dargestellte Analyse macht anhand von Posts einiger AfD-Mitglieder deutlich, dass Politiker Online-Plattformen wie Twitter/X nicht nur nutzen, um direkt auf die Öffentlichkeit einzugehen und einen Austausch zu fördern, sondern auch, um im Diskurs miteinander zu kommunizieren. So etablieren sich auf Twitter/X bestimmte Kreise, die den Politikern eine Bühne bieten. Durch die Diskreditierung des Gegners und die Kommentierung eines bestimmten Ereignisses nutzen die AfD-Mitglieder Twitter/X als Plattform zur Darstellung ihrer Agenda und ihrer selbst. Vermutlich erhoffen sie sich von ihren Posts eine Wirkung auf das aktuelle politische Panorama und damit die Durchsetzung ihrer politischen Ziele. Im Gegensatz zur traditionellen externen politischen Kommunikation, bei der Politiker Pressekonferenzen einberufen und Fernsehauftritte planen müssen, haben sie bei Posts die Möglichkeit, ihre Meinung unmittelbar zu äußern und die Reaktionen darauf zeitnah zu lesen. Posts fungieren somit als eine Art direkte Meinungsumfrage für Politiker.

6. Fazit

Alle aufgezeigten Posts belegen die ablehnende Haltung von AfD-Politikern gegenüber ihren Gegnern sowie der nationalen bzw. internationalen Politik Deutschlands. Es handelt sich vor allem um sprecherzentrierte Sprechakte, die die Meinung des jeweiligen Politikers über ein bestimmtes Thema ausdrücken. Zudem sind es sachverhaltszentrierte Sprechakte, da sie eine Handlung mit einer Wertung des Sprechers verbinden. Schließlich können sie auch als gegnerbezogene Sprechakte gelten, da sie ein emotionales Verhalten des jeweiligen Followers gegenüber einem in der konkreten Interaktion anwesenden oder abwesenden politischen Gegner darstellen.

Bei der Untersuchung zeigten sich ferner folgende Aspekte: Auf Twitter herrschen eine Neigung zur Behandlung brandaktueller Themen und eine Tendenz zur Dialoghaftigkeit vor. Die AfD-Parteimitglieder, die ihre Posts veröffentlichen, streben de facto die Interaktion mit ihren Followern an. Weiter ist die Genauigkeit der lexikalischen Auswahl zu betonen, die mit der festgelegten Kürze bei Posts verbunden ist. Diese Posts weisen zudem – mit Ausnahme derjenigen von Weidel und Höcke (s. Bsp. 7, 8, 9) – eine Sprache auf, die in erster Linie Merkmale der Mündlichkeit aufweist.

Ferner wird klar, dass Twitter/X als soziales Medium wesentlich zur Vermittlung von politischen Themen beiträgt, einerseits weil die Politiker dadurch ein Massenpublikum erreichen können, mit dem sie unmittelbar interagieren; andererseits erlaubt es den Bürgern, sich ein Bild der aktuellen politischen Situation zu machen, Informationen über die drängendsten gesellschaftlichen Probleme zu bekommen usw. Durch die Kürze der Posts vermitteln die Politiker nicht nur ihre Ideen und die Ideologie ihrer Partei, sondern sie personalisieren damit auch ihr öffentliches Image und gestalten ihre mediale Agenda.

Die vorgestellte Analyse bietet einen wertvollen Einblick in die politische Online-Kommunikation einer spezifischen deutschen Partei. Die Schnelligkeit und Gleichzeitigkeit, die das Medium Twitter/X bietet, erlauben es, dynamische Veränderungen zeitnah zu erfassen. Um das Verständnis weiter zu vertiefen, wäre es bereichernd, die Untersuchungen auf andere deutsche Parteien sowie kontrastiv auf andere Sprachen, wie beispielsweise Italienisch, auszuweiten.

Bionote: Isabella Ferron holds a master degree in Foreign Languages and Literatures and a master degree in Linguistics from the Ca' Foscari University in Venice; she obtained her PhD at the LMU Munich in a joint study program with the University of Padua. Thanks to various scholarships (DAAD, DLA – German Literatur Archive Marbach, Institute of Modern Languages, London, University of Potsdam) she has continued her academic education in an international context. She has collaborated with several universities as a

lecturer (Venice, Padua, Siena, Milan and Leipzig) and has been a postdoctoral fellow at the Ca' Foscari University of Venice and at the Italian Institute of Germanic Studies (Rome). She was RTDA (junior researcher) of German Language and Translation at the Department of Political Science of Sapienza University and is currently RTDB (senior researcher) of German Language and Translation at the Department of Linguistic and Cultural Studies of the University of Modena and Reggio Emilia. Her academic interests lie in the fields of language history, political language, contrastive (Italian-German) and cognitive linguistics, multilingualism and the relationship between language and literature.

Authors' address: isabella.ferron@unimore.it

Literatur

- Austin J. 1962, *How to do Things with Words*, Cambridge University Press, Cambridge.
- Baecker D. 2017, *Wie verändert die Digitalisierung unser Denken und unseren Umgang mit der Welt? Ausgangspunkte einer Theorie der Digitalisierung*, in Gläß R. und Leukert B. (Hrsg.), *Handel 4.0: Die Digitalisierung des Handels – Strategien, Technologien, Transformation*, Springer, Berlin/Heidelberg, S. 3-24.
- Beck K. 2020, *Kommunikationswissenschaft*, (6. Aufl.), UVK, München.
- Beißwenger M. und Storrer A. 2012, *Interaktionsorientiertes Schreiben und interaktive Lesespiele in der Chat-Kommunikation*, in „Zeitschrift für Literaturwissenschaft und Linguistik“ 168, S. 92-124.
- Benkler Y. 2006, *The Wealth of Networks*, Yale University Press, New Haven.
- Bruns A. and Highfield T. 2016, *Is Habermas on Twitter? Social Media and the Public Sphere*, in Enli G., Bruns A., Larsson A.O., Skogerbo E. and Christensen C. (eds.), *The Routledge Companion to Social Media and Politics*, Routledge, New York/London, S. 56-73.
- Busemann K. und Gscheidle C. 2009, *Web2.0: Communitys bei jungen Nutzern beliebt*, in „Media Perspektive“ 7, S. 356-364.
- Crystal D. 2005, *The Scope of Internet Linguistics. Paper given online to the American Association for the Advancement of Science Meeting, Februar 2005*, verfügbar unter <https://davidcrystal.com/Files/BooksAndArticles/-4113.pdf>, (letzter Zugriff: 30.01.2024).
- Deppermann A. 2018, *Sprache in der multimodalen Interaktion*, in Deppermann A. und Reineke S. (Hrsg.), *Sprache im kommunikativen, interaktiven und kulturellen Kontext*, De Gruyter, Berlin, S. 51-87.
- Digitales Wörterbuch der deutschen Sprache (DWDS)*, verfügbar unter <https://www.dwds.de>, (letzter Zugriff: 07.07.2024).
- Eilders C. 2006, *News factors and news decisions. Theoretical and methodological advances in Germany*, in „Communication“ 31 [1], S. 5-24.
- Emmer M., Vowe G. und Wollig J. 2011, *Bürger online: Die Entwicklung der politischen Kommunikation in Deutschland*, UVK, Konstanz.
- Feezell J.T. 2018, *Agenda setting through social media: The importance of incidental news exposure and social filtering in the digital era*, in „Political Research Quarterly“ 71 [2], S. 482-494.
- Finkbeiner R., Meibauer J. und Wiese H. 2016, *What is perjoration and how can it be expressed in language?*, in Finkbeiner R., Meibauer J. und Wiese H. (eds.), *Pejoration*, Benjamin, Amsterdam, S. 1-18.
- Fuchs C. 2021, *Das Digitale Kapital. Zur Kritik der politischen Ökonomie des 21. Jahrhunderts*, mandelbaum kritik & utopie, Wien/ Berlin.
- Gilardi F., Gessler T., Kubli M. and Müller S. 2022, *Social Media and Political Agenda Setting*, in „Political Communication“ 39 [1], S. 39-60.
- Gilbert M.A. 2001, *Emotional Messages*, in „Argumentation. An International Journal on Reasoning“ 15, S. 239-249.
- Girnth H. 2013, *Sprache und Sprachverwendung in der Politik. Eine Einführung in die linguistische Analyse öffentlich-politischer Kommunikation*, De Gruyter, Berlin.
- Haase M., Huber M., Krumeich A. und Rehm G. 1997, *Internetkommunikation und Sprachwandel*, in Weingarten R. (Hrsg.), *Sprachwandel durch Computer*, Westdeutscher Verlag, Opladen, S. 51-85.
- Herring S.C. and Honeycutt C. 2009, *Beyond. Microblogging: Conversation and Collaboration via Twitter*, in „Proceedings of the Forty-Second Hawai'i International

- Conference on System Sciences (HICSS-42)", Los Alamitos, CA. IEEE Press, 2009, verfügbar unter <http://ella.slis.indiana.edu/~herring/honeycutt.herring.2009.pdf>, (letzter Zugriff 23.02.2024).
- Holly W. 2011, *Medien, Kommunikationsformen, Textsortenfamilien*, in Habscheid S. (Hrsg.), *Textsorten, Handlungsmuster, Oberflächen. Linguistische Typologien der Kommunikation*, De Gruyter, Berlin, S. 144-163.
- Jarren O. und Donges P. 2002, *Politische Kommunikation in der Mediengesellschaft. Eine Einführung*, Springer VS, Wiesbaden.
- Jarren O. und Sarcinelli U. 1998, „Politische Kommunikation“ als Forschungs- und als politisches Handlungsfeld: Einleitende Anmerkungen zum Versuch der systematischen Erschließung, in Jarren O., Sarcinelli U. und Saxer U. (Hrsg.), *Politische Kommunikation in der demokratischen Gesellschaft. Ein Handbuch mit Lexikonteil*, Westdeutscher Verlag, Opladen, S. 13-20.
- Java A., Song X., Finin T. and Tseng B. 2007, *Why We Twitter: Understanding Microblogging Usage and Communities*, in "Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop", S. 56-65.
- Koch P. und Oesterreicher W. 1985, *Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte*, in Holtus G., Metzeltin M. und Schmitt C. (Hrsg.), *Lexikon der Romanischen Linguistik (LRL)*, Band / Vol I.2., Niemeyer, Tübingen, S. 15-43.
- Koch P. und Oesterreicher W. 2008, *Mündlichkeit und Schriftlichkeit von Texten*, in Janich N. (Hrsg.), *Textlinguistik. 15 Einführungen*, Narr, Tübingen, S. 199-215.
- Korte K.-R. und Hirscher G. (Hrsg.) 2000, *Darstellungspolitik oder Entscheidungspolitik? Über den Wandel von Politikstilen in westlichen Demokratien*, Publikationen der Hanns-Seidel-Stiftung, München.
- Krämer S. 2001, *Sprache, Sprechakt. Kommunikation. Sprachtheoretische Positionen des 20. Jahrhunderts*, Suhrkamp, Frankfurt a.M.
- Krotz F. 2001, *Die Mediatisierung kommunikativen Handelns. Wie sich Alltag und soziale Beziehungen, Kultur und Gesellschaft durch die Medien wandeln*, Westdeutscher Verlag, Opladen.
- Krotz F. 2007, *Mediatisierung. Fallstudien zum Wandel von Kommunikation*, Westdeutscher Verlag, Opladen.
- Luhmann N. 1984, *Soziale Systeme. Grundriß einer allgemeinen Theorie*, Suhrkamp, Frankfurt a.M.
- Mair P. and Katz R.S. 1995, *Changing Models of Party Organization and Party Democracy. The Emergency of the Cartel Party*, in "Party Politics" 1 [1], S. 5-28.
- Marcinkowski F. (Hrsg.) 2014, *Framing als politischer Prozess. Beiträge zum Deutungskampf in der politischen Kommunikation*, Nomos, Baden-Baden.
- Moraldo S. 2009, *Twitter: Kommunikationsplattform zwischen Nachrichtendienst, Small Talk und SMS*, in Moraldo S. (Hrsg.), *Internet.kom. Neue Sprach- und Kommunikationsformen im WorldWideWeb*, Band 1: *Kommunikationsformen*, Aracne, Rom, S. 245-281.
- Moraldo S. 2012a, „das Leben in 140 Zeichen... heisst Twitter :-)" " Teil I: Kommunikative Aspekte der Microblogging-Plattform Twitter, in „Sprachspiegel“ 3, S. 77-85.
- Moraldo S. 2012b, „das Leben in 140 Zeichen... heisst Twitter :-)" " Teil II: Sprachliche Aspekte der Microblogging-Plattform Twitter, in „Sprachspiegel“ 4, S. 2-13.
- Moraldo S. 2012c, "Obwohl Korrektur: Polizei Hat Gebäude im coolen Duisburger Innenhafen". Die Kommunikationsplattform Twitter an der Schnittstelle zwischen Sprechsprachlichkeit und medial bedingter Schriftlichkeit, in Güthner S., Imo W., Meer

- D. und Schneider J.G. (Hrsg.), *Kommunikation und Öffentlichkeit. Sprachwissenschaftliche Potenziale zwischen Empirie und Norm*, De Gruyter, Berlin, S. 183-208.
- Murthy D. 2013, *Twitter. Social Communication in the Twitter Age*, Polity, Cambridge.
- Niehr T. 2007, *Schlagwort*, in Ueding G. (Hg.), *Historisches Wörterbuch der Rhetorik*, Bd. 8, De Gruyter, Berlin, S. 496-502.
- Niehr T. 2021, *Politischer Sprachgebrauch*, in "Lublin Studies in Modern Languages and Literature" 45 [1], S. 75-85.
- O'Reilly T. und Milstein S. 2009, *Das Twitter-Buch*, O'Reilly, Köln.
- Runkel J., Schlobinski P. und Torsten S. 1998, *Sprache und Kommunikation im Internet*, in „Muttersprache. Vierteljahresschrift für deutsche Sprache“ 2, S. 97-109.
- Sarcinelli U. 2009, *Politische Kommunikation. in Deutschland. Zur Politikvermittlung im demokratischen System*, Springer VS, Wiesbaden.
- Searle J.R. 1981, *Expression and Meaning. Studies in the Theory of Speech Acts*, Cambridge University Press, London/New York.
- Stark B., Magin M. und Jürgens P. 2021, *Maßlos überschätzt. Ein Überblick über theoretische Annahmen und empirische Befunde zu Filterblasen und Echokammern*, in Eisenegger M., Prinzing M., Ettinger P. und Blum R. (Hrsg.), *Digitaler Strukturwandel der Öffentlichkeit. Mediensymposium*, Springer VS, Wiesbaden, S. 303-312.
- Stiegler C. 2015, *Digitale Medientheorien*, in Stiegler C., Breitenbach P. und Zorbach T. (Hrsg.), *New Media Culture: Mediale Phänomene der Netzkultur*, transcript, Bielefeld, S. 11-28.
- Storrer A. 2009, *Rhetorisch-stilistische Eigenschaften der Sprache des Internets*, in Fix U. (Hrsg.), *Rhetorik und Stilistik. Ein internationales Handbuch historischer und systematischer Forschung*, Bd. 2, De Gruyter, Berlin, S. 2211-2226.
- Technau B. 2018, *Beleidigungswörter. Die Semantik und Pragmatik pejorativer Personenbezeichnungen*, De Gruyter, Berlin.
- Thaler V. 2003, *Chat-Kommunikation im Spannungsfeld zwischen Oralität und Literalität*, VWF, Berlin.
- Thimm C., Dang-Anh M. und Einspänner J. 2011, *Diskurssystem Twitter: Semiotische und handlungstheoretische Perspektiven*, in Anastasiadis M. und Thimm C. (Hrsg.), *Social Media. Theorie und Praxis digitaler Sozialität*, Peter Lang, Frankfurt a.M., S. 265-285.
- Vowe G. 2020, *Digitalisierung als grundlegender Veränderungsprozess der politischen Kommunikation*, in Borucki I., Kleinen-von Königslöw K., Marschall S. und Zerback T. (Hrsg.), *Handbuch Politische Kommunikation*, Springer VS, Wiesbaden, S. 1-15.
- Wolton D. 1990, *Political Communication: The Construction of a Model*, in "European Journal of Communication" 5 [1], S. 9-28.

LES MÉTAPHORES NO-VAX SUR TWITTER

Une étude comparative français-italien

GLORIA ZANELLA

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Abstract – The aim of this research is to study metaphors related to the topic of anti-vaccination used by users in digital communication, especially on Twitter¹. We created a bilingual comparable corpus in French and Italian based on a collection of tweets characterised by no-vax metaphors that were published online between March 2020 and October 2022. The qualitative analysis of the tweets considered shows the use of a variety of metaphors in order to emphasise two different positions taken by users on Twitter: on the one hand, the conspiracy, the fight for freedom, sacrifice, being victims of a health system that forces people to be vaccinated, resistance against the health dictatorship and against the restrictions imposed by the government; on the other hand, the free choice to be vaccinated, the fight against the virus and against the circulation of fake news. This study shows the role of metaphors, which are used to insult or denigrate, as well as to defend and support one's own position in relation to the vaccination against the Covid-19 virus.

Keywords: comparable corpus; Covid-19; digital communication; no-vax metaphors; Twitter.

1. Introduction

Les épidémies précédentes comme la grippe porcine (Mundwiler 2013), la grippe aviaire (Moirand 2007 ; Koteyko *et al.* 2008), le syndrome respiratoire aigu sévère (SRAS) (Wallis, Nerlich 2005), Ebola (Balteiro 2017) et le virus Zika (Ribeiro *et al.* 2018) n'ont pas été caractérisées par une médiatisation très intense dans les réseaux sociaux, par rapport à la pandémie de Covid-19. Après les cas de virus 2019-nCoV attestés à Wuhan en Chine, en février 2020, aussi dans les régions italiennes de Lombardie et de Vénétie nous avons assisté à la confirmation des premiers cas d'infection et de mort liés au virus Covid-19. En mars 2020, le président du Conseil des ministres italien Giuseppe Conte annonçait l'institution des « zones rouges » renforcées², des « zones rouges »,

¹ Dans cette contribution, nous parlons de Twitter et le corpus comparable qui fait l'objet des analyses est constitué de tweets car au moment du développement de cette recherche, la plateforme de microblogage n'était pas encore rebaptisée X par Elon Musk.

² Consulter le site du portail officiel du ministère de la Santé italien : <https://www.salute.gov.it/portale/nuovocoronavirus/dettaglioNotizieNuovoCoronavirus.jsp?id=4137>

des « zones orange » et des « zones jaunes » pour établir les différents niveaux de la contagion et déclarer le confinement et la quarantaine, d'abord seulement dans certaines communes et régions³, ensuite dans tout le territoire national⁴. À partir de ces premiers événements vécus en Italie, la pandémie de Covid-19 a profondément bouleversé non seulement notre vie quotidienne, mais aussi les dynamiques discursives et ses effets se sont manifestés aussi dans les réseaux socio-numériques tels que Twitter. En outre, la médiatisation extrême qui a caractérisé la crise sanitaire causée par le virus Covid-19 a vu la circulation d'une énorme quantité de fausses nouvelles qui ont contribué au tourbillon de l'*infodémie*. Ce mot-valise, créé par l'OMS en 2020⁵, indique la circulation d'un excès d'informations erronées qui menace la gestion de la pandémie de Covid-19 et qui contribue à la méfiance à l'égard des autorités sanitaires⁶.

Dans cette étude, nous nous penchons sur des textes brefs publiés sur Twitter, à savoir les tweets⁷, qui sont caractérisés par les métaphores nommées *no-vax* : ces métaphores sont produites par les usagers qui s'opposent à la vaccination contre le virus Covid-19 ou par les internautes qui s'adressent à eux. Dans l'*Enciclopedia Italiana Treccani*, le terme *no-vax* est entré dans la section des néologismes en 2017 et sa définition indique : « Chi, che è contrario alla vaccinazione e, in particolare, al fatto di sottoporre la popolazione infantile alla profilassi vaccinale »⁸. Pour ce qui concerne le français, la définition d'« antivax » proposée par le dictionnaire *Larousse* est :

Se dit d'un mouvement d'opinion marqué par une opposition à certains vaccins ou à la vaccination en général, dont il remet en cause l'efficacité et l'innocuité. (On dit aussi antivaccin, antivaccination ou antivaccinal.) Les théories de ce mouvement, non fondées sur les données acquises de la science, exposent, dans la mesure où elles entraînent une baisse de la couverture vaccinale, à un risque de réapparition d'épidémies de maladies infectieuses.⁹

³ À ce propos, nous conseillons de consulter le site Web qui concerne la Gazzetta ufficiale della Repubblica italiana : <https://www.gazzettaufficiale.it/eli/id/2020/03/08/20A01522/sg>

⁴ Voir le site du portail officiel du ministère de la Santé italien : <https://www.salute.gov.it/portale/nuovocoronavirus/dettaglioNotizieNuovoCoronavirus.jsp?id=4184>

⁵ Nous conseillons de se rendre sur le site de l'Organisation Mondiale de la Santé : <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation>

⁶ Voir le site Web de l'Organisation Mondiale de la Santé : https://www.who.int/health-topics/infodemic#tab=tab_1

⁷ Les tweets sont des textes caractérisés par une longueur maximale de 280 caractères qui sont publiés sur le réseau social Twitter par les utilisateurs.

⁸ Consulter la définition proposée sur le site du dictionnaire Treccani : https://www.treccani.it/vocabolario/no-vax_res-79d2489e-89da-11e8-a7cb-00271042e8d9_%28Neologismi%29/

⁹ À ce propos, nous conseillons de voir la définition proposée sur le site du dictionnaire *Larousse* : <https://www.larousse.fr/dictionnaires/francais/antivax/188487>

Cette contribution propose une analyse qualitative qui concerne les métaphores *no-vax* comme mécanisme discursif puissant dans le cadrage de l'opinion. Ces métaphores peuplent les tweets tirés d'un corpus comparable français-italien constitué de deux cents tweets publiés. Le corpus qui fait l'objet de cette étude a été constitué entre mars 2020 et octobre 2022 et les tweets extraits sont aussi caractérisés par la présence des mots-dièse : #novax #covid #Covid19. Dans la deuxième section de cette contribution, nous présentons le cadre théorique à la base de cette étude qui a permis d'orienter la recherche concernant, d'un côté, le rôle joué par les métaphores dans la communication numérique ; de l'autre côté, les métaphores *no-vax* qui peuplent la twittosphère à travers leur créativité. Dans la troisième section, nous présentons la méthodologie utilisée et les analyses des métaphores créatives qui caractérisent les tweets extraits du corpus comparable français et italien. Enfin, dans la quatrième section, nous proposons les remarques et les réflexions conclusives de cette étude.

2. Les métaphores dans la communication numérique

La métaphore est considérée comme un outil essentiel dans la communication et dans la pensée, à partir de la notion de métaphore conceptuelle et du cadre théorique donnés par Lakoff et Johnson (1985) et grâce à l'intégration de la théorie du *conceptual blending* proposée par Fauconnier et Turner (2002). Dans l'ouvrage fondateur *Les métaphores dans la vie quotidienne*, Lakoff et Johnson affirment que : « L'essence d'une métaphore est qu'elle permet de comprendre quelque chose (et d'en faire l'expérience) en termes de quelque chose d'autre » (1985, p. 15). Plus récemment, Anderson et Semino (2020) mettent en évidence l'importance du rôle joué par la métaphore dans le domaine des humanités numériques : à ce propos, elles affirment que les environnements numériques peuvent favoriser la création de nouveaux modèles dans l'utilisation des métaphores. En outre, ces environnements numériques peuvent être façonnés par les métaphores sélectionnées pour des buts différents. Selon les auteurs, la métaphore représente un sujet intéressant dans les humanités numériques pour deux raisons : premièrement, les facultés propres de l'être humain comme parler, penser et comprendre une chose en termes d'une autre permettent de développer les dimensions de la communication, du raisonnement et de l'imagination ; deuxièmement, les métaphores deviennent aussi une façon d'exprimer ce qui est proprement humain dans différents domaines : de la vie quotidienne à la littérature, des chansons à l'art visuel. En particulier, Anderson (2020) a examiné la métaphore dans l'ère numérique et le domaine sémantique de l'eau en mouvement, à savoir les marées, les vagues et les inondations, grâce à l'utilisation de « Metaphor Map of English » à partir des données de l'*Historical Thesaurus of English*.

Dans la communication numérique, les informations données par les tweets publiés sur le réseau socio-numérique Twitter représentent un réservoir intéressant pour l'analyse du discours (Paveau 2017). Bon nombre de recherches liées au discours dans les réseaux sociaux concernent différentes thématiques comme la communication dans les situations d'urgence et de danger telles que les catastrophes et les calamités naturelles, mais aussi les attaques dénigrantes à travers Twitter. Yeo *et al.* (2022) ont proposé une étude basée sur la communication sur Twitter à propos de l'inondation en Louisiane du Sud qui a eu lieu en 2016. Cette recherche a démontré la dynamique des modèles de communication dans les réseaux socio-numériques pendant la reprise à long terme afin d'identifier les préoccupations vécues par les communautés touchées par cet événement et de comprendre les sentiments des individus. Selon Yeo *et al.* : « Social media provides an interactive and direct platform for emergency managers to initiate and manage two-way risk communication and implement proper strategies to influence risk perception and behaviors of people » (2022, p. 1682). Un autre exemple est représenté par la recherche de Martín *et al.* (2017) dont les auteurs ont examiné le discours sur Twitter concernant l'ouragan Matthew à travers une analyse spatiotemporelle. La thématique liée aux attaques dénigrantes sur Twitter a été examinée par Demjén et Hardaker (2017) qui ont proposé une recherche sur les métaphores offensives utilisées comme stratégies discursives afin d'attaquer la journaliste britannique Caroline Criado-Perez pendant la campagne d'inclusion de l'image d'une figure historique féminine sur les billets de banque anglais.

Pendant la pandémie de Covid-19, la communication a joué un rôle clé pour la connaissance du public à propos de la situation d'émergence et la circulation des mesures sanitaires comme le confinement, la quarantaine, la distance sociale et les gestes barrière afin de contenir la diffusion du virus et d'adopter des comportements préventifs pour la santé (Sabat *et al.* 2020). En particulier, le discours en ligne sur Twitter a fait l'objet des recherches empiriques menées par Wicke et Bolognesi sur l'utilisation des métaphores pendant la pandémie de Covid-19 et les auteurs affirment que :

Online discourse on Twitter [...] has recently attracted a number of contributions, because the texts (the tweets) found on this platform are considered to be a good proxy for the public opinion and perception related to the pandemic that we are currently experiencing (Bruns and Weller, 2016). It follows that understanding and interpreting such discourse, its evolution over time, and its interdependence with real-world events can help us understand how people conceptualize and react to the global crisis. (Wicke, Bolognesi 2021)

En particulier, dans leurs analyses, Wicke et Bolognesi (2020, 2021) se focalisent sur les thématiques abordées sur les tweets qui concernent la pandémie de Covid-19 et ils remarquent la présence de métaphores liées au

cadre de la guerre comme « conflit », « lutter », « combat », « bataille » et l'identification du virus comme « ennemi » à vaincre. Toutefois, une résistance aux métaphores guerrières utilisées pendant la pandémie de Covid-19 telles que « soldats », « bataille » et « première ligne » a fait l'objet de discussions dans de nombreux débats non seulement publics, mais aussi académiques (Bates 2020 ; Sabucedo *et al.* 2020 ; Semino 2021a ; Serhan 2020). En effet, le cadre métaphorique de la lutte contre le virus associe la survivance au combat, mais, en même temps, il semble que les victimes décédées à cause de la pandémie de Covid-19 n'aient pas lutté suffisamment. Pour sa part, Rossi (2022) explore les concepts clés de la pandémie de Covid-19 et l'auteure remarque que la métaphore guerrière est privilégiée par l'italien, mais le français est orienté vers la métaphore fluide de l'inondation. La perception des différents cadres métaphoriques utilisés pendant la pandémie de Covid-19 en Allemagne, en Italie et aux Pays-Bas a fait l'objet d'une recherche menée par Brugman *et al.* (2022) et dans les résultats, les descriptions métaphoriques montrent une différence entre les domaines sources et les contextes nationaux.

Pour ce qui concerne les métaphores utilisées contre l'hésitation à la vaccination du virus Covid-19, Semino (2021b) examine deux métaphores telles que *cooking roast dinners* et *wearing seatbelts in cars* qui apparaissent dans l'ouvrage *Vaxxers* (Gilbert, Green 2021) afin d'expliquer la rapidité de développement du vaccin et la protection qu'il offre contre le virus Covid-19. Dans le cas de la métaphore *cooking roast dinners*, les « ingrédients » de base sont mis ensemble pour produire un vaccin en mesure de faire face à l'agent pathogène spécifique comme le SARS-CoV-2. Dans la métaphore *wearing seatbelts in cars*, le *mapping* est basé sur la sécurité : les ceintures de sécurité ne sont pas en mesure d'éviter un accident, mais elles protègent des blessures les plus profondes, ainsi que le vaccin n'empêche pas l'infection par le virus, mais il représente une protection contre les symptômes les plus graves.

Dans le IX chapitre « Magic, Miracle Cures and Metaphoric Thought in the Anti-Vaccine Movement », Charteris-Black (2021) s'intéresse, d'abord, aux remèdes miracles qui circulaient pendant la pandémie de Covid-19, comme l'utilisation de la chloroquine et de l'hydroxychloroquine proposée par Donald Trump aux États-Unis et Jair Bolsonaro au Brésil ; ensuite, l'auteur se focalise sur les théories complotistes qui affirmaient que Bill Gates avait créé une micropuce afin de surveiller la population et que le 5G était la cause de la diffusion du virus. Les théories complotistes liées aux risques du vaccin sont souvent ancrées aux sentiments de méfiance envers les autorités. En particulier, Charteris-Black remarque que l'Autorité n'est pas acceptée comme moralement juste par tous et quelqu'un peut s'opposer au contrôle de l'état sur son corps, comme dans le cas des anti-vaccins. À propos des concepts d'autorité et de réputation dans les réseaux sociaux, Vicari affirme que :

Dans le web 2.0, la notion d'autorité s'enrichit d'une dimension technique. Pour faire autorité, il ne suffit pas d'être populaires, à savoir de recevoir un grand nombre de clics ou de *likes*, il faut aussi construire progressivement et dans le temps sa propre réputation. [...] Du point de vue technique, la mesure de l'autorité d'une page web serait ainsi déterminée par la présence sur d'autres sites de liens hypertextuels qui pointent vers la page elle-même. Ces algorithmes introduisent donc un indice de confiance extrait des pratiques hypertextuelles des usagers et permettent d'augmenter la réputation des résultats considérés comme plus pertinents. (Vicari 2021)

Dans le web 2.0, autorité, visibilité, popularité et réputation non seulement se confondent du moins partiellement, mais elles acquièrent aussi une dimension technique : si l'autorité d'une page web ou d'un profil mesure le nombre de liens sur d'autres sites pointant cette page ou profil, la réputation mesurerait plutôt la pertinence de ces liens. D'une conception de l'autorité plus « verticale », venant d'en haut, le web 2.0 favoriserait du point de vue technique et algorithmique aussi, des processus de confiance dans le discours autrui se situant plutôt sur l'axe « horizontal » ainsi qu'une certaine fluidité des relations d'autorité, dont l'étude ne peut pas faire l'impasse de la nature « technodiscursive » (Paveau 2017) de toute production en ligne. (Vicari 2022)

Enfin, le rôle de premier plan joué par les métaphores créatives et conflictuelles (Prandi 2012, 2017, 2021) contribue à développer l'intérêt pour l'étude de la responsabilité et de l'effacement énonciatif (Koren 2006 ; Moirand 2006 ; Rabatel 2004a, 2004b) dans la communication numérique (Paveau 2016 ; Vicari 2023) et de la circulation de fausses nouvelles liées à la pandémie Covid-19 (Farroni 2021 ; Brachotte *et al.* 2022 ; Giry, Nouvel 2022). Dans la section suivante, nous proposons la méthodologie adoptée et la présentation de l'analyse qualitative des tweets qui font l'objet de cette étude.

3. Méthodologie et analyse qualitative des métaphores *no-vax* sur Twitter

La méthodologie choisie pour cette étude se penche sur la procédure pour l'identification des métaphores dans la linguistique de corpus, à savoir la MIP (en français PIM, Procédure pour l'Identification de Métaphores) proposée par Pragglejaz Group (2007) et puis développée par Steen *et al.* (2010) comme MIPVU. Cette procédure nous permet de déterminer l'usage métaphorique des unités lexicales à travers l'établissement du sens de base et la validation du sens contextuel afin de remarquer si le sens contextuel s'oppose au sens de base, mais il peut être compris en comparaison avec ce dernier. En particulier, nous avons appliqué la PIM au français comme proposé par Reijnierse (2019), considérant les dictionnaires *Le Petit Robert*, *Le Trésor de la Langue Française informatisé* (TLFi) et *Larousse*, pour le corpus français ; *Treccani*, *Garzanti* et *Nuovo De Mauro*, pour le corpus italien. En outre, la théorie de la métaphore conceptuelle (Lakoff, Johnson 1985 ; Kövecses 2010) et la théorie

de la métaphore conceptuelle étendue (Kövecses 2020) ont été prises en considération comme fondement théorique pour l'analyse qui fait l'objet de cette étude. Pour ce qui concerne le rôle des métaphores *no-vax* dans le corpus constitué, nous nous intéressons aussi à la notion de contextualisation, proposée par Kövecses :

As regards contextualization, the extended view involves four types of context and the contextual factors they subsume: Situational context: physical, social, and cultural; Discourse context: surrounding discourse, previous discourse, knowledge about main elements of discourse, and dominant forms of discourse; Bodily context: correlations in experience, bodily conditions, and body specificities; Conceptual-cognitive context: metaphorical conceptual system, ideology, concerns and interests, and history. (Kövecses 2020, p. 168)

Dans cette étude, la procédure d'identification des métaphores (PIM) a été appliquée manuellement afin d'établir la métaphoricité de chaque unité lexicale et de mener une analyse qualitative sur les métaphores *no-vax* qui caractérisent les tweets dans le corpus comparable bilingue français-italien. En particulier, les critères à la base de la sélection des tweets qui font l'objet de l'étude sont trois :

1. critère chronologique : les tweets ont été publiés par les utilisateurs Twitter entre mars 2020 et octobre 2022 ;
2. critère thématique : le discours *no-vax* caractérisé par les métaphores dans les tweets publiés par les usagers ;
3. la présence de mots-dièse dans les tweets, à savoir : #novax, #covid, #Covid19.

Une ultérieure remarque sur le corpus comparable bilingue français-italien concerne la présence des mots-dièse qui caractérisent les tweets considérés, tels que :

Corpus français	Corpus italien
#novax	#novax
#Défenseurs des #Libertés	#libertà
#JusticePourLesSuspendus	#complotto
#DictatureSanitaire	#Dittaturasanitaria
#JeNeSuisPlusCitoyen	#DittaturaNazisanitaria
#NonAuPassDeLaHonte	#Norimberga2
#NonAuPassSanitaire	#NoGreenPass
#NonAuPassVaccinal	#SuperGreenPass
#NonAuVaccinObligatoire	#covididiots
#Apartheidsanitaire	#disinformazione

Tableau 1
Mots-dièse dans le corpus comparable français-italien.

Dans cette contribution, notre intérêt est focalisé sur l'analyse qualitative de huit tweets extraits du corpus français et de neuf tweets tirés du corpus italien afin d'explorer dans le détail les métaphores créatives présentes dans les tweets publiés par les usagers sur la thématique *no-vax* dans la twittosphère entre 2020 et 2022.

3.1. Les métaphores no-vax dans les tweets extraits du corpus français

Pendant la pandémie de Covid-19, la métaphore a démontré son importance capitale, non seulement dans le tissu social et anthropologique, mais aussi dans le contexte discursif. À partir de la constitution du corpus français, nous proposons une analyse qualitative de huit tweets extraits afin d'examiner les cas les plus intéressants sur les métaphores no-vax produites par les citoyens-internautes. Dans le premier tweet que nous considérons, nous pouvons voir que le « pass sanitaire » est la cause d'un système sanitaire à deux vitesses qui fait des différences parmi les patients hospitalisés et qui ne soigne pas ceux qui ne sont pas vaccinés. De cette façon, les anti-vaccins sont des victimes forcées à la vaccination obligatoire afin d'avoir accès aux soins hospitaliers :

- (1) #Italie : Refus de #Soins, pour cette dame âgée, qui vient pour un suivi #Diabétique. Le #PassSanitaire, créé un #Système de #Santé à 2 vitesses. Les #NoVax sont poussés au #VaccinationObligatoire, pour accéder à l'#Hôpital. En #France, même chose, mais ça n'émeut presque pas. (19/10/2021)

La métaphore créative à la base de ce tweet permet de conceptualiser le système sanitaire comme une organisation qui avance selon deux vitesses différentes : les personnes munies du pass sanitaire peuvent avoir accès plus rapidement aux soins tandis que ceux qui ne l'ont pas sont des individus mis de côté. Dans ce tweet, nous pouvons remarquer que la prise de parole de l'utilisateur s'oriente vers l'anonymat, toutefois la circulation de ces nouvelles contre le pass sanitaire s'enracine de plus en plus.

La métaphore conceptuelle LES NOVAX SONT DES DÉFENSEURS DES LIBERTÉS¹⁰ est présente dans ce tweet et elle structure le domaine lié au mouvement anti-vax :

- (2) Encore une aberration, qui devrait arriver en #France, d'ici peu et qui ne choquera que les #NoVax #Défenseurs des #Libertés... (24/10/2021)

Comme nous pouvons constater, dans le tweet (2), les anti-vaccins deviennent les défenseurs des libertés qui luttent pendant une période caractérisée par la campagne de vaccination. Bon nombre d'exemples du corpus sont enracinés dans la métaphore conceptuelle qui considère les anti-vaccins comme les

¹⁰ Pour les interactions conceptuelles nous adopterons le protocole de citation en lettres majuscules qui en usage pour les modèles cognitifs idéalisés (MCI) dans le sillage de Lakoff et Johnson.

défenseurs des libertés qui combattent la guerre contre la vaccination obligatoire et le système sanitaire qui l'impose.

Dans l'exemple suivant (3), nous examinons la métaphore conceptuelle LE SYSTÈME SANITAIRE EST UNE DICTATURE. Dans cette dictature sanitaire, il n'y a pas le libre choix de ne pas adhérer à la vaccination. Pour cette raison, la vaccination devient une obligation imposée par un régime dictatorial. Cette « perte de chance » pour le vaccin représente le risque de tomber malade et de devenir les responsables de la transmission du virus aux autres. En outre, le mot-dièse #JeNeSuisPlusCitoyen montre la perte des droits du citoyen et le processus de déshumanisation subi à cause de la dictature sanitaire et le pass sanitaire devient le « pass de la honte » :

- (3) Mise en danger de la vie d'autrui par « perte de chance »... #DictatureSanitaire
#JeNeSuisPlusCitoyen #NonAuPassDeLaHonte #NonAuPassSanitaire
#NonAuPassVaccinal #NonAuVaccinObligatoire #Covid_19 #COVID19
(07/10/2022)

La métaphore permet de transférer le concept de la dictature du domaine politique au domaine sanitaire. Le lien créé par la métaphore est immédiat : le régime dictatorial caractérisé par l'imposition despotique et intransigeante de l'autorité passe de la politique au domaine sanitaire. Nous pouvons voir que cette métaphore est présente aussi dans le tweet qui suit (4) dont l'utilisateur exprime la soumission de la population au totalitarisme covidiste et à la vaccination obligatoire.

- (4) Les français vont-ils se réveiller un jour ? Cette soumission volontaire de la population au totalitarisme covidiste est effrayante. #COVID19 #novax
#AntiPassSanitaire (16/09/2021)

Les deux exemples qui suivent (5, 6) se focalisent sur un autre déclenchement généré par la métaphore : l'apartheid citoyen remarque la séparation entre vaccinés et non-vaccinés créée dans la société pendant la pandémie de Covid-19 et à cause de l'imposition du « pass sanitaire » les libertés se trouvent dans l'impasse :

- (5) Honte à tous ces politiques professionnels de l'enrichissement personnel, de la corruption #BigPharma, de l'apartheid citoyen. Non, la reprise épidémique n'est pas du fait des #novax, elle est du fait de leur incompétence à miser tout, sur un #vaccin #Pfiser qui n'en est pas un apartheid sanitaire. Le « pass sanitaire », nos libertés dans l'impasse. (05/12/2021)

Dans le tweet (6), nous remarquons la responsabilité énonciative par l'usage du *je* et la métaphore conflictuelle de l'apartheid causée par les antivaccins passe du domaine sanitaire à la politique. En outre, la démocratie devient un être vivant qui est tué par les choix politiques :

- (6) Je vous ai écouté la à l instant le #MichelTaube : c est quoi ça de dire que les #novax ont créé l #Apartheid en #France ? Stop Mensonges : C est #Macron & co qui ont coupé la #France en 2 & tué la #democratie ! Avec 1 taux 90% de #vacciner, laissez donc les #novax tranquilles (31/12/2021)

Dans le tweet (7), l'utilisateur s'adresse directement aux anti-vaccins. À travers le conflit déclenché par la métaphore, le manque de liberté vécu par les anti-vaccins à cause du « pass sanitaire » est rapproché au manque d'air éprouvé par le malade de Covid-19 qui doit utiliser le respirateur pour survivre :

- (7) Aux : #NO : #novax #NoMasks #nocovid #Nolockdowns #antivax
« Le #Vaccin restreint votre #liberté ? essayez le #respirateur... » Ext. 11 : La peste : « La bêtise insiste toujours, on s'en apercevrait si l'on ne pensait pas toujours à soi » (Camus) (01/11/2021)

Nous pouvons constater que le ressort heuristique de la métaphore permet de rapprocher le concept du vaccin en tant que violation de la liberté de choix pour les anti-vaccins du respirateur utilisé pour soigner les malades de Covid-19 qui représente les restrictions liées au manque d'air, un symptôme sérieux causé par le virus. Cette métaphore conflictuelle est double : le vaccin restreint la liberté à cause de son imposition, tandis que le respirateur réduit la liberté des malades à cause des effets du virus. En outre, ce tweet se clôt sur une citation tirée du roman *La Peste* d'Albert Camus.

Le dernier tweet (8) extrait du corpus français est adressé aux anti-vaccins qui contribuent à l'infodémie et à la circulation de fausses nouvelles sur Twitter, puisque le vaccin devient le liquide diabolique qui menace la santé de la population :

- (8) Pour les #novax : ceux qui hurlent en font des dégâts en écrivant sur #Twitter :
« Ils jettent quelque chose dans le #Vaccin : le liquide de #Satan » (17/01/2022)

Le conflit déclenché par le transfert du concept du vaccin au domaine de la religion représente un exemple de métaphore conflictuelle qui s'ouvre à plusieurs interprétations : le vaccin est un liquide diabolique, à considérer avec méfiance puisque sa composition et ses effets sont inconnus ; le vaccin au lieu de protéger du risque d'une maladie, il représente un ennemi qui menace la santé des individus. En outre, le « liquide de Satan » est un cas de métaphore isolée dans le corpus analysé.

Les exemples considérés à partir du corpus français montrent une stratification de métaphores : les anti-vaccins sont les victimes du système sanitaire et les défenseurs des libertés combattants contre la dictature sanitaire et le totalitarisme covidiste qui imposent la vaccination. À cause du refus du vaccin, les anti-vaccins sont les coupables et les responsables de la diffusion du virus. En outre, le vaccin est le liquide de Satan, il crée un système à deux vitesses qui fait des différences dans le soin des patients et le « pass sanitaire » est la cause de l'apartheid dans la société pendant la pandémie de Covid-19.

Dans les tweets analysés, nous pouvons remarquer une dilution de la responsabilité énonciative des usagers anti-vaccins : le tweet est un moyen de partage d'informations avec la communauté des internautes dans la twittosphère. Même si les sources et l'authenticité de ces informations ne sont pas vérifiées, leur contenu continue à circuler dans les tweets et à s'enraciner dans l'opinion publique. Les anti-vax adoptent une posture de combattants contre le vaccin, de défenseurs des libertés et de porte-paroles qui luttent contre la dictature sanitaire, mais ils sont aussi les victimes d'un apartheid sanitaire. Enfin, Twitter devient un réseau caractérisé par des métaphores conflictuelles et bon nombre de métaphores no-vax sont liées au domaine source du combat et de la lutte contre la dictature sanitaire.

3.2. Les tweets du corpus italien et les métaphores no-vax

Les tweets extraits du corpus italien qui font l'objet de cette analyse qualitative sont neuf. Dans le premier tweet, l'utilisateur affirme que les vaccinés ont une horloge dedans et le temps de vie est donné par rapport à la typologie de vaccin qu'ils ont reçu :

- (1) Tutti i sierati hanno un "orologio" a tempo dentro di sé, le cui lancette girano a seconda del lotto di veleno che gli è capitato. Ma x tutti il tempo scadrà, entro pochi anni. I più fortunati max 10 anni. Sapete come funziona il veleno per topi? Non uccide subito se no gli altri topi ne accorgono (07/10/2022)

Dans ce tweet, la métaphore porte sur le temps de vie des vaccinés qui est donné par les aiguilles de cette horloge qui tournent selon le lot de poison que les personnes ont reçu. De cette façon, nous assistons à une projection : d'un côté, la durée de vie symbolisée par l'horloge qui marque le temps ; de l'autre côté, le vaccin associé au poison. En outre, la vie des vaccinés est en danger puisque le vaccin est comme le poison à rat qui tue lentement.

Dans le tweet suivant (2), l'utilisateur s'adresse au gouvernement et il soutient que les données sur la contagion sont manipulées :

- (2) Dichiarate dati e risultati manipolati per un pubblico di spaventati creduloni! Siete la vergogna del genere umano e i vostri sieri killer sono la distruzione della salute e dei diritti fondamentali dell'uomo. #ViVi #novax #NoGreenPass (10/03/2022)

En particulier, dans ce tweet (2), nous rencontrons un exemple de métaphore conflictuelle dont le vaccin est un sérum tueur qui, au lieu d'offrir une protection, est la cause de la destruction de la santé et des droits fondamentaux de l'homme.

Un exemple de métaphore guerrière est représenté par la lutte pour la liberté et les droits violés :

- (3) La lotta per la Libertà e Diritti di TUTTI si esaurirà solo quando si inizierà a rispettarli! Fino a quel momento scriveremo, grideremo e diffonderemo la Verità sempre più OVUNQUE! #ViVi #DittaturaNazisanitaria #novax #Norimberga2 (10/03/2022)

Dans ce tweet (3), l'usager utilise la métaphore de la guerre pour déclarer que les anti-vaccins sont les combattants pour la liberté et les droits. En outre, il est intéressant de remarquer les mots-dièse utilisés #DittaturaNazisanitaria et #Norimberga2. Ces mots-dièse font référence à une dictature qualifiée comme « nazie sanitaire » et imposée par le gouvernement italien et aux lois de Nuremberg qui renvoient à la persécution.

Les exemples qui suivent (4, 5, 6, 7) nous pouvons rencontrer l'usage de la première personne au singulier et du *nous* qui remarque la construction d'un nouvel *ethos* collectif. En outre, les tweets sont caractérisés par deux métaphores dominantes : premièrement, les anti-vaccins considérés comme les victimes d'un système sanitaire qui impose la vaccination ; deuxièmement, la lutte contre un système qui discrimine la population et menace la paix des individus :

- (4) Io comunque non potrò mai perdonare tutto ciò che ci hanno fatto subire in questi anni, le offese, le discriminazioni, le difficoltà che ci hanno fatto affrontare. Per me non vi sarà nessun perdono e nessuna riconciliazione. #Covid_19 #SuperGreenPass #greenpass #novax (17/03/2022)
- (5) Dobbiamo insegnarlo ai nostri figli di lottare sempre. Anche a costo di essere emarginati (17/03/2022)
- (6) E non dobbiamo dimenticare! Ci hanno rubato il sonno, la pace e 2 anni della nostra mia vita... (17/03/2022)
- (7) Ci risentiamo dopo l'estate, quando rimetteranno il greenpass con obbligo di quarta dose. Non siamo più persone, ma lucine verdi e rosse che possono essere accese e spente a loro piacimento, e chi non ubbidisce è fuori. Poi vediamo quali cocci di dignità si raccoglieranno. (20/03/2022)

Dans le tweet (7), nous rencontrons une métaphore conflictuelle : la vaccination obligatoire constitue une privation de liberté, mais elle réduit aussi la population à des petites lumières vertes et rouges qui peuvent être allumées ou éteintes par un système sanitaire qui exclut les individus n'obéissant pas aux règles. En outre, la dignité de l'être humain est cassée en morceaux, puisque les individus perdent leur identité.

Dans l'exemple qui suit (8), l'usager utilise la première personne du singulier et il partage son opinion :

- (8) Al di là del fatto che sarò soddisfatto solo quando abrogheranno l'infame tessera verde per legge, oltreché la museruola e obbligo di inoculazione di un farmaco sperimentale, la democrazia non è solo una forma, ma soprattutto una sostanza, è il rispetto delle persone che sono chiamate a votare. La disinvoltura con la quale si è penetrati nella sfera privata in questi due anni, il controllo sugli

spostamenti, ha superato ampiamente il lecito e le reali necessità dell'emergenza e ha fatto scempio di questo rispetto, e ora la democrazia. (17/03/2022)

En particulier, dans ce tweet le « green pass » devient « l'infâme carte verte » ; la muselière et la vaccination obligatoire d'un médicament expérimental représentent la perte de la liberté des individus et la démocratie est mise en danger à cause du contrôle de la population et de la vie privée.

Le dernier tweet (9) a été publié quand en Italie l'accès aux lieux publics était interdit aux personnes non-vaccinées :

(9) Dal #1aprile per i #novax SCATTA poi LA MODALITÀ CANE, per bar, risto e locali? Fuori sì ma dentro no? Voi andrete? (18/03/2021)

La métaphore conflictuelle qui concerne la « modalità cane » fait référence aux personnes dépourvues du « pass sanitaire » qui pouvaient s'installer seulement dehors des locaux, comme le chien qui attend son propriétaire dehors du magasin ou du café. La projection créée par cette métaphore est forte : les non-vaccinés perdent leur dimension humaine et ils doivent rester à l'extérieur comme les chiens auxquels l'accès n'est pas autorisé.

À partir des exemples extraits du corpus italien, nous pouvons remarquer l'usage du tweet pour partager ses opinions en matière de Covid-19 et de vaccination. Le corpus est caractérisé par la présence de la métaphore guerrière qui est utilisée pour exprimer la lutte pour la liberté et les droits dans la démocratie qui sont menacés par le système sanitaire. Deux métaphores récurrentes dans le corpus italien concernent : le vaccin qui est associé à un poison et à un sérum tueur représentant une menace pour la vie des individus ; les anti-vaccins qui sont des victimes d'un système qui les persécute. Enfin, les individus qui s'opposent à la vaccination perdent leur dimension humaine : d'un côté, ils deviennent des petites lumières qui peuvent être allumées ou éteintes par les décisions du gouvernement ; de l'autre côté, ils sont des chiens avec la muselière et auxquels l'accès aux lieux publics est interdit.

4. En guise de conclusion

Dans les tweets qui ont fait l'objet de cette étude comparative français-italien, nous avons vu la capacité de cadrage des métaphores *no-vax* sur Twitter. Un éventail de métaphores créatives et délibérément conflictuelles a été exploité par les usagers dans la twittosphère afin de discuter du virus et du vaccin, mais aussi pour diffuser de fausses nouvelles en contribuant au chaos informatif en matière de vaccination et de mesures sanitaires. Bon nombre d'exemples considérés remarquent la présence de métaphores guerrières qui ont été utilisées pour les « combats » contre le totalitarisme covidiste, les « défenseurs » de la liberté mise en danger, les « luttes » pour les droits de

l'homme menacés par la dictature et l'apartheid sanitaires pendant la pandémie de Covid-19.

Dans les corpus français et italien analysés, les métaphores *no-vax* sont un outil puissant de cadrage conceptuel et un moyen de partage d'opinions parmi les usagers. Les métaphores *no-vax* se focalisent sur les anti-vaccins comme défenseurs de la liberté, mais aussi ils sont les victimes de la dictature sanitaire, d'un « système à deux vitesses » et de la persécution pendant la pandémie de Covid-19. Nous remarquons aussi que la métaphore conceptuelle LE VACCIN EST UN POISON est partagée entre les deux corpus et elle déclenche de nouveaux transferts : le vaccin n'offre pas de protection, en revanche il menace la santé, il devient « le liquide de Satan », un poison à rats et un sérum tueur. L'analyse du corpus comparable bilingue français-italien montre aussi des différences : la métaphore qui concerne l'apartheid sanitaire a été rencontrée seulement dans le corpus français et la métaphore « modalità cane », faisant référence aux anti-vaccins qui sans le « pass sanitaire » n'ont pas accès aux lieux publics, est un cas de métaphore isolée puisqu'elle est présente seulement dans le corpus italien. Les métaphores dans les discours *no-vax* deviennent ainsi des stratégies offensives qui expriment les différents sentiments et opinions des internautes dans la twittosphère. Les tweets analysés dans le corpus français sont caractérisés par une dilution de la responsabilité énonciative des usagers anti-vaccins et par une stratification énonciative, puisque les informations continuent leur circulation à l'intérieur de la twittosphère et elles s'enracinent de plus en plus, même si les sources ne sont pas vérifiées. En particulier, le corpus italien est caractérisé par l'emploi du *nous*, ce qui permet d'enraciner l'idée de collectivité et de groupe uni qui partage un point de vue commun. Enfin, l'étude des métaphores sur les réseaux socio-numériques est susceptible d'ultérieures recherches puisque ces environnements numériques donnent la possibilité d'avoir accès à l'opinion des usagers et du public à propos de thématiques sociales et d'actualité.

Bionote : Gloria Zanella est doctorante en Sciences humaines au Département d'Études linguistiques et culturelles à l'Université de Modène et de Reggio d'Émilie. Elle développe son projet de recherche sur les métaphores dans l'astrophysique en diachronie courte, sous la direction des professeures Chiara Preite de l'Université de Modène et de Reggio d'Émilie et Aurélie Picton de l'Université de Genève. Ses intérêts de recherche se concentrent sur la métaphore et la variation diachronique courte dans la vulgarisation scientifique et dans les terminologies des domaines technico-scientifiques. Sa première contribution intitulée « Astrophysique et vulgarisation. Une étude de cas » a été publiée dans les *Quaderni del CIRM* (2021). Elle a aussi mené des recherches sur les métaphores dans le domaine de la boulangerie en français et en italien.

Adresse de l'auteure : gloria.zanella@unimore.it

Références bibliographiques

- Anderson W. 2020, *Metaphor in the digital age: Opening the flood-gates*, in “Token: A Journal of English Linguistics” 10, pp. 15-38.
- Anderson W. and Semino E. 2020, *Metaphor*, in Adolph S. and Knight D. (eds.), *The Routledge Handbook of English Language and Digital Humanities*, Routledge, London/New York, pp. 125-142.
- Balteiro I. 2017, *Metaphor in Ebola’s popularized scientific discourse*, in “Iberica” 34, pp. 209-230.
- Bates B.R. 2020, *The (in)appropriateness of the WAR metaphor in response to SARS-CoV-2: A rapid analysis of Donald J. Trump’s rhetoric*, in “Frontiers in Communication” 5, pp. 1-12.
- Brachotte G., Frame A., Gautier L., Nazarov W. et Selmi A. 2022, *Les discours complotistes sur Twitter à propos de la vaccination contre la Covid-19 en France : communautés et analyse sémio-linguistique des hashtags*, in « Mots. Les langages du politique » 130, pp. 79-103.
- Brugman B.C., Droog E., Reijniere W.G., Leymann S., Frezza G. and Renardel de Lavalette K.Y. 2022, *Audience Perceptions of COVID-19 Metaphors: The Role of Source Domain and Country Context*, in “Metaphor and Symbol” 37 [2], pp. 101-113.
- Bruns A. and Weller K. 2016, *Twitter as a first draft of the present: and the challenges of preserving it for the future*, in *Proceedings of the 8th ACM Conference on Web Science (Hannover)*, pp. 183-189.
- Charteris-Black J. 2021, *Metaphors of Coronavirus: Invisible Enemy or Zombie Apocalypse?*, Springer Nature. [doi:10.1007/978-3-030-85106-4](https://doi.org/10.1007/978-3-030-85106-4)
- Demjén Z. and Hardaker C. 2017, *Metaphor, impoliteness, and offence in online communication*, in Semino E. and Demjén Z. (eds.), *The Routledge Handbook of Metaphor and Language*, Routledge, London/New York, pp. 353-367.
- Farroni C. 2021, *L’era del vaccino anti-Covid tra bufale e Aluhüte. Analisi comparativa del lessico italiano e tedesco tramite l’utilizzo di corpora*, in *Parole che non c’erano. La lingua e le lingue nel contesto della pandemia*, Pollicino S. e Zanot I. (a cura di), Roma TrE-Press, Rome, pp. 153-168.
- Fauconnier G. and Turner M. 2002, *The Way We Think: Conceptual Blending and the Mind’s Hidden Complexities*, Basic Books, New York.
- Gilbert S. and Green C. 2021, *Vaxxers*, Hodder & Stoughton, London.
- Giry J. et Nouvel D. 2022, *Étudier les discours « conspirationnistes » et leur circulation sur Twitter. Les théories du complot comme objets du traitement automatique du langage et de l’analyse des données textuelles*, in « Mots. Les langages du politique » 130, pp. 37-55.
- Koren R. 2006, *La responsabilité des Uns dans le regard des Autres : l’effacement énonciatif au prisme de la prise de position argumentative*, in « Semen » 22. doi.org/10.4000/semen.2820
- Koteyko N., Brown B. and Crawford P. 2008, *The dead parrot and the dying swan: The role of metaphor scenarios in UK press coverage of avian flu in the UK in 2005-2006*, in “Metaphor and Symbol” 23 [4], pp. 242-261.
- Kövecses Z. 2010, *Metaphor. A Practical Introduction* (second edition), Oxford University Press, Oxford.
- Kövecses Z. 2020, *Extended conceptual metaphor theory*, Cambridge University Press, Cambridge.

- Lakoff G. et Johnson M. 1985, *Les métaphores dans la vie quotidienne*, Éd. de Minuit, Paris.
- Martín Y., Li Z. and Cutter S.L. 2017, *Leveraging Twitter to gauge evacuation compliance: spatiotemporal analysis of Hurricane Matthew*, in “PLoS ONE” 12 [7]. [10.1371/journal.pone.0181701](https://doi.org/10.1371/journal.pone.0181701)
- Moirand S. 2006, *Responsabilité et énonciation dans la presse quotidienne : questionnements sur les observables et les catégories d’analyse*, in « Semen » 22. doi.org/10.4000/semen.2798
- Moirand S. 2007, *Les discours de la presse quotidienne. Observer, analyser, comprendre*, Presses Universitaires de France, Paris.
- Mundwiler V. 2013, ‘CATCH IT, BIN IT, KILL IT’ - *On the metaphorical conceptualisation of the 2009 swine flu pandemic in British media text*, in “Metaphorik.de” 24, pp. 37-63.
- Paveau M.-A. 2016, *Éthique du discours numérique*, in “Línguas e Instrumentos Línguísticos” 37, pp. 177-210.
- Paveau M.-A. 2017, *L’analyse du discours numérique. Dictionnaire des formes et des pratiques*, Hermann, Paris.
- Pragglejaz Group 2007, *MIP: a method for identifying metaphorically used words in discourse*, in “Metaphor and Symbol” 22 [1], pp. 1-39.
- Prandi M. 2012, *A Plea for Living Metaphors: Conflictual Metaphors and Metaphorical Swarms*, in “Metaphor and Symbol” 27 [2], pp. 148-170.
- Prandi M. 2017, *Conceptual Conflicts in Metaphors and Figurative Language*, Routledge, New York/London.
- Prandi M. 2021, *Le metafore tra le figure: una mappa ragionata*, UTET Università, Turin.
- Rabatel A. 2004a, *Effacement énonciatif et effets argumentatifs indirects dans l’incipit du Mort qu’il faut de Semprun*, in « Semen » 17. doi.org/10.4000/semen.2334
- Rabatel A. 2004b, *L’effacement énonciatif dans les discours rapportés et ses effets pragmatiques*, in « Langages » 156 [4], pp. 3-17.
- Reijnierse W.G. 2019, *Linguistic metaphor identification in French*, in Nacey S., Dorst A. G., Krennmayr T. and Reijnierse W.G. (ed.), *Metaphor Identification in Multiple Languages*, John Benjamins, Amsterdam/Philadelphia, pp. 69-90.
- Ribeiro B., Hartley S., Nerlich B. and Jaspal R. 2018, *Media coverage of the Zika crisis in Brazil: The construction of a “war” frame that masked social and gender inequalities*, in “Social Science & Medicine” 200, pp. 137-144.
- Rossi M. 2022, *Vulgariser les concepts scientifiques dans la presse : les définitions par métaphore*, in « Roczniki Humanistyczne » LXX [8], pp. 77-94.
- Sabat I., Neumann-Böhme S., Varghese N.E., Barros P.P., Brouwer W., van Exel J., Schreyögg J. and Stargardt T. 2020, *United but divided: Policy responses and people’s perceptions in the EU during the COVID-19 outbreak*, in “Health Policy” 124 [9], pp. 909-918.
- Sabucedo J.-M., Alzate M. and Hur D. 2020, *COVID-19 and the metaphor of war (COVID-19 y la metáfora de la guerra)*, in “International Journal of Social Psychology” 35 [3], pp. 618-624.
- Semino E. 2021a, *Not Soldiers but Fire-fighters - Metaphors and Covid-19*, in “Health Communication” 36 [1], pp. 50-58.
- Semino E. 2021b, *From roast dinners to seatbelts: Metaphors to address Covid-19 vaccine hesitancy*, in “Questioning Vaccination Discourse”, <https://www.lancaster.ac.uk/vaccination-discourse/2021/09/12/from-roast-dinners-to-seatbelts-metaphors-to-address-covid-19-vaccine-hesitancy-by-elena-semino/> (10.09.2024)

- Serhan Y. 2020, *The Case Against Waging 'War' on the Coronavirus*, in "The Atlantic", <https://www.theatlantic.com/international/archive/2020/03/war-metaphor-coronavirus/609049/> (10.09.2024)
- Steen G.J., Dorst A.G., Herrmann J.B., Kaal A.A., Krennmayr T. and Pasma T. 2010, *A Method for Linguistic Metaphor Identification. From MIP to MIPVU*, John Benjamins, Amsterdam.
- Vicari S. 2021, *Introduction : Autorité et Web 2.0*, in « Argumentation et Analyse du Discours » 26. doi.org/10.4000/aad.4936
- Vicari S. 2022, *Quand les médecins deviennent influenceurs : la vulgarisation des termes de la Covid-19 dans Facebook, Instagram et Twitter*, in "Repères DoRiF" 25 – Le lexique de la pandémie et ses variantes, <https://www.dorif.it/reperes/stefano-vicari-quand-les-medecins-deviennent-influenceurs-la-vulgarisation-des-termes-de-la-covid-19-dans-facebook-instagram-et-twitter/> (Consulté le 10.09.2024)
- Vicari S. 2023, *Discours d'influenceurs, discours d'autorité ? Le cas de deux médecins influenceurs sur Twitter*, in « Argumentation et Analyse du Discours » 30. doi.org/10.4000/aad.7505
- Wallis P. and Nerlich B. 2005, *Disease metaphors in new epidemics: The UK media framing of the 2003 SARS epidemic*, in "Social Science & Medicine" 60 [11], pp. 2629-2639.
- Wicke P. and Bolognesi M.M. 2021, *Covid-19 Discourse on Twitter: How the Topics, Sentiments, Subjectivity, and Figurative Frames Changed Over Time*, in "Frontiers in Communication" 6. doi.org/10.3389/fcomm.2021.651997
- Wicke P. and Bolognesi M.M. 2020, *Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter*, in "PLoS ONE" 15 [9]. doi.org/10.1371/journal.pone.0240010
- Yeo J., Knox C.C. and Hu Q. 2022, *Disaster Recovery Communication in the Digital Era: Social Media and the 2016 Southern Louisiana Flood*, in "Risk Analysis" 42 [8], pp. 1670-1685.

© 2024 University of Salento - Coordinamento SIBA



<http://siba.unisalento.it>