Lingue e Linguaggi
Lingue Linguaggi 56 (2023), 53-68
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v56p53
http://siba-ese.unisalento.it, © 2023 Università del Salento

SOCIOLINGUISTIC VARIATIONIST ANALYSIS OF WORD-EMOTION LEXICON IN COOK ISLANDS ENGLISH ONLINE NEWS

CARMEN CIANCIA, SIRIA GUZZO[1]
UNIVERSITÀ DEGLI STUDI DI SALERNO

Abstract – This paper describes how journalists, in the Cook Islands, use sentiment lexicon when reporting online news. To do so, we employ Sentiment Analysis (SA) in combination with sociolinguistic variationist theory and logistic regression analysis. SA relies on the *Word-Emotion Association Lexicon* source (Mohammad, Turney 2013), which comprises 10,170 lexical items. The bulk of research carried out on sentiment analysis only distinguishes between positive vs. negative emotions. By contrast, we provide a fine-grained coding by exploring how eight specific core emotions (i.e. ANGER, ANTICIPATION, FEAR, DISGUST, JOY, SADNESS, SURPRISE, and TRUST) are socially stratified in formal contexts. We built a small-scale corpus from web-based newspapers to find out (i) whether social factors (age and sex) condition the use of sentiment lexicon and (ii) to evaluate the socially acknowledged generalisations according to which females tend to use sentiment lexicon more than males. The data was quantitatively examined through mixed-effects Rbrul logistic regression analysis. The independent variables include: word class (i.e. nous, adjectives, verbs), sex, age, and word-frequency. Specifically, the latter is a variable involved in language processing and is commonly studied in psycholinguistics, sociolinguistics, and corpus linguistics (Mickiewicz 2019). To account for word-frequency we use the SUBTLEX-US corpus (Brysbaert, New 2009). Our findings suggest that sentiment lexicon is conditioned by age, with young and old speakers favouring the use of sentiment lexicon. Sex, word class, and word-frequency do not have a significant influence on sentiment lexicon in our data.

Keywords: Sentiment Analysis; Word-Emotion Lexicon; Cook Islands English; Online News.

1. Introduction

Language enables people to interact with one another, to communicate emotions, and to develop emotions into specific categories. Sociolinguistic studies have widely documented the interrelation between language and society at the phonological, lexical, morphosyntactic levels (Holmes, Meyerhoff 2008) as well as in terms of social identity. In detail, micro-sociolinguistic studies examine how the social structure affects the way people talk and how patterns of use are influenced by social factors (e.g. social class, age, and sex). Macro-sociolinguistic studies investigate what societies do with their languages, such as language shift, maintenance, attitudes "which account for the functional distribution of speech forms in society" (Coulmas 1997, p. 2). Sociolinguistic research has also devoted attention to the influence of social factors on language variation in online communication such as public discussions (O'Connor *et al.* 2010), in blog posts, Twitter,

[1] The content of this paper was designed, investigated, and co-written by both authors. Siria Guzzo is responsible for the Abstract, Sections 1, 2 and 3, and Carmen Ciancia is responsible for Sections 4, 5, 6 and 7.

instant messages (Koch *et al.* 2022), but what remains unexplored is the relationship between language, social factors, and emotions. The relationship between language and emotions has been explored in different research fields, such as psychology, anthropology, linguistics, and neuroscience. However, little corpus-based sociolinguistic research on how individuals convey their emotions in formal contexts has been carried out (Schweinberger 2019).

To examine the emotions contained in a text, most researchers use Sentiment Analysis,[2] a growing field at the intersection of linguistics and computer science. Sentiment analysis is commonly used to extract information from (a) positive and negative words within a text, (b) from the context of words, as well as (c) from the structure of a text. Sentiment analysis is also employed in social media analysis conducted by marketers, in behaviour analyses (Pak, Paroubek 2010; Weller *et al.* 2014); political discourses (Bakliwal *et al.* 2013); and it is used to find out whether online reviews are positive or negative.

Our present study adopts Sentiment Analysis to investigate to what extent emotionality is conveyed through lexicon in the Cook Islands English online news, as little research has been carried out on L2 varieties of English in the South Pacific. Moreover, since anthropological studies have revealed that the degree of emotional expressivity, verbally and non-verbally, varies from culture to culture, it would be interesting to find out how emotionality is expressed in post-colonial contexts. This led us to the following research questions: do social factors (e.g. sex, and age) play a significant role in the way sentiment lexicon is used when delivering online news? Is it true, as socially acknowledged, that females are more likely to use sentiment lexicon more than males? The influence of speakers' sex in language use has long been accounted for in sociolinguistics, starting from the early 1970s. Previous sociolinguistic studies (Labov 1990) suggest that men tend to use non-standard variants more than women, whereas in the linguistic changes below the level of social awareness, women tend to use the incoming variants more than men. Besides sex, the age of individuals is of notable importance, in variationist studies, as it represents "a place in history and a life stage" (Eckert 1997, p. 151). When linguistic variables are stratified by age, they reflect a change in the community. Thus, we wonder whether emotionality can be also significantly stratified by sex and age. Additionally, sociolinguistic studies have also demonstrated that linguistic variables are not only correlated with age, sex, and class of individuals, but also with style. Results show that non-standard variants are commonly used in spontaneous speech, whereas standard variants were found to be adopted across all social classes as formality increases. Style, as a sociolinguistic variable, was first examined in terms of attention paid to speech (Labov 1972). This means that, when speaking, people change their styles mostly in response to how much attention they are giving to the speech itself. Sociolinguistic studies, indeed, have found that people speak differently on different occasions. A student, for instance, speaks in a certain way when talking to his supervisor, but speaks very differently (e.g. uses non-standard words, non-standard sounds, etc.) when chatting to his friends. This study focuses on formality (intended as attention paid to speech) where the authors are expected to pay more attention to the words they use when delivering/writing online news. We hypothesize that in formal contexts a high level of

[2] The term sentiment analysis is commonly adopted to refer to approaches which extract information on emotion from natural language (Crossley et al. 2016; Hutto and Gilbert 2014). In this paper, however, we use the terms *emotion* and *sentiment* interchangeably, but we distinguish them in the literature review section.

non-emotional words might be employed. Therefore, our study aims at finding out whether sentiment lexicon, in the context of news delivering, shows the use of non-emotional words over emotional ones, or viceversa.

Let us now provide a brief overview of the socio-linguistic situation of the Cook Islands, before turning the attention to the methods adopted.

2. A brief overlook at the socio-linguistic situation of the Cook Islands

The Cook Islands is a descriptive name given to 15 tiny islands spreading over 2 million square kilometers of the South Pacific Ocean divided into two distinct groups: the Southern Cook Islands and the Northern Cook Islands where Rarotonga is the capital island. No studies have concentrated on the language(s) of the Cook Islands only and the English spoken on the islands has not been investigated and given specific single attention so far.

Previous studies have concentrated on the post-colonial variety of English known as “South Pacific English” or “Pasifika” English(es) which is a relatively new field of study and included the non-native varieties of English spoken in Fiji, Samoa and the Cook Islands, among others (Bell, Gibson 2008; Kachru 1982, 1985, 1992, 2006; Schneider 2003, 2007). Most of the studies have looked at Pasifika English in the language of Pacific Island migrants to New Zealand as showing signs of mutual accommodation and homogenization with NZE although the ethnic island origins remain recognizably distinct (Kachru, Nelson 2006). Similarities as well as differences in those varieties due, among other reasons, to the Melanesian and Polynesian substrate influence do exist. Due to its geographical and political closeness, New Zealand English may replace, in some of the Cook Islands, the former prestigious American and British varieties and become the new model for the national standard. More recently, Carolin Bewer (2008, 2012) field researched in the Cook Islands, Samoa and Fiji and compared their grammatical structure and their sociolinguistic profile, looking at concord patterns in South Pacific Englishes, and paying strong attention to the influence of New Zealand English and the local substrate. Interesting results emerged in terms of modals and semi modals of obligation and necessity and most of her conclusions led to the existence of an acrolectal English in the South Pacific which stands as an emerging standard.

The languages spoken in those 15 islands include English, Cook Islands Māori, Rarotongan, Pukapukan, and other local dialects. In this context the educational system has shown to have had an enormous impact on the development of language attitudes among Cook Islands speakers. Today it is a local mother tongue which is predestined to take over the function of nation building and the choice of one language: one of the Maori languages – over English, or the other way round, becomes a political issue.In the Cook Islands, English is mostly spoken as a L2, except for the capital Rarotonga where it is spoken as an L1. Different ‘-lects’ of English can be distinguished according to social class, regional upbringing and the type of schooling: the acrolects are determined by the educated elite in the urban centre of Rarotonga.

The Cook Islands have always had a particularly close relationship with New Zealand as they are traditionally and politically associated with New Zealand and a close cooperation with New Zealand about education has been long-standing. English is the dominant language in literature and the media. English is the language of the workplace and business, as in most L2 varieties of English in post-colonial environments.

Significantly, the English acrolect in the Cook Islands is shaped by exonormative forces (i.e., New Zealand English or American English arriving through Samoan English) as well as by tendencies towards nativisation in a constant language contact situation (Bewer 2012). The variety of English spoken in the Cook Islands, and more specifically in Rarotonga, is clearly influenced by the local Maori substrates (Guzzo forthcoming).
Being so remote from a European perspective, they are socially isolated and strongly holding on to their native cultures. Some of the most remote islands suffer from an even higher level of isolation which in some cases leads to Limited English Proficient (LEP). Apart from those living in the most tourist locations, most of the islanders are socioeconomically disadvantaged, poor and have received very little formal education (Bewer 2012). Their condition has proven to be very difficult on some of the most remote islands, which in some periods can be isolated for very long periods due to difficult weather conditions and the impossibility for cargos to bring provisions. High levels of migration, mainly towards New Zealand, have affected these areas and entire generations of young Cook Islanders have moved and settled down elsewhere looking for education, more opportunities, and a better living. All of this has led us to some research questions which link sentiment analysis and Cook Islands English, in a variationist sociolinguistic framework.

3. Previous Studies on Emotion, and Sentiment Analysis

From a psychological perspective, it is often assumed that emotion categories hold a universal essence. Similarly, research on the emotion lexicon argues that language is a representation of emotion categories which already exists (Lindquist, Barrett 2008b). Whatever the theoretical background, there is no doubt that words may evoke different emotions in different contexts. They frequently convey affect and they do so either explicitly via their core meaning (e.g. denotation) or implicitly via connotations. For instance, the word *dejected* denotes sadness, whereas *failure* connotates sadness – thus, they are both associated to the same emotion.[3]

Previous studies suggest that the expression of emotional states, or affect is classified in two main categories: one category expresses judgment towards people, whereas a second category expresses appreciation, or aesthetic opinion (Martin & White 2005). The combination of these two categories shows how individuals convey feelings. How people express emotions is analysed under different umbrella terms in Linguistics and in other social sciences. Research in Linguistics and Psychology have examined how individuals express, understand, and are influenced by the expression of subjectivity in terms of positive or negative opinions (Krippendorf 2004). Batson *et al.* (1992) focused on the role of affect, Traugott (2010) examined subjectivity and point of view, Aikhenvald (2004) investigated evidentiality, Biber and Finegan (1989) looked at attitudinal stance, Portner (2009) analysed modality, Martin and White (2005) examined appraisal to find out how people use language to convey emotion, evaluation, and subjectivity. The solely theorical interest in the investigation of subjectivity and evaluation, in the last few years, has been associated with an increased attention towards individuals' opinion expressed on the internet.

[3] A collection of words and emotions which signals specific emotions is referred to as *word-affect association lexicon* (aka *emotion lexicon*, aka *sentiment lexicon*). Throughout this paper, we adopt the term *sentiment lexicon*.

What has received the most attention are polarity lexicon and the semantic orientation of words (i.e. their positive and negative values). A number of lexicon-based approaches went beyond the binary division of words, which denotes their positivity and negativity, and included information related to strength. This means that a word can be positive - strong (e.g. *altruistic*), positive – weak (e.g. *accept*), neutral (e.g. *alliance*), negative – weak (e.g. *addiction*), and negative – strong (e.g. *abuse*) (Wiebe *et al.* 2004).

The bulk of sentiment research has focused on texts, whereas studies on emotions usually have examined speech in relation to suprasegmental features (e.g. prosody, pitch, and intonation). English is usually the target language for examining sentiment, yet some studies have also attempted to investigate how sentiment is expressed in other languages, such as Arabic (El-Beltagy, Ali 2013; Salameh *et al.* 2015), Chinese (Huang *et al.* 2012; Wan 2008; Wang *et al.* 2012), French (Benamara *et al.* 2013; Ghorbel 2012; Marchand 2012), German (Haas, Versley2015; Waltinger 2010), and Spanish (López *et al.* 2012; Molina-González *et al.* 2013; Moreno-Ortiz, Pérez Hernández 2012; Vilares *et al.* 2015).[4]

Wierzbicka (1999) states that languages differ in the way they express emotions. The Germans distinguish between *Eifersucht* (a relation with somebody else) and *Neid* (related to material possession), whereas the Dutch only use one word – *jaloezie* (“jealous, envy”) – for both terms. The Greek do not seem to have a term which expresses frustration (Pavlenko 2008), whilst Dholuo - an African language - with the term *maof* indicates “the feeling of desiring to see relatives and friends that have not been seen for too long and is by extension transferred to other things” (Omondi 1997, p. 97).The question as to why the above differences between languages influence the way people speak or experience their feelings is beyond the purpose of this study (see Colombetti, 2009; Lindquist 2008 for further details).

Most of sentiment analysis has focused on the evaluative function of adjectives. The subjective information in a text is often conveyed by adjectives, indeed, much work has gone into determining the adjectives' semantic orientation. According to the Pollyanna Principle (Boucher, Osgood 1969), opinions expressed online reveal a high frequency of positive adjectives, and other positive words. In this respect, the literature suggests that a great deal of online review (e.g. TripAdvisor’s reviews) are mainly positive because, according to Jing-Schmidt (2007), people tend to use less negative words due to political correctness. Commonly, negative words or statements appear to be perceived as more marked than positive ones, both pragmatically and psychologically (Horn 1989). Boucher and Osgood (1969) claim that positive terms outnumber negative ones as individuals tend to remember past events in a positive way. Foolen (2012, p. 351) claims that nouns like *love, anger, surprise* index emotions as well as other parts of speech such as verbs (*to love, to fear*, etc.), adjectives (*happy, sad,* etc.), and prepositions (e.g. *love for something*) which play a notable role “in the relational aspects of the conceptualization of emotions”.

Indeed, more recent studies of sentiment analysis have noticed that besides the function of adjectives, sentiment is also expressed through other parts of speech, namely nouns, verbs, adverbs (Benamara *et al.* 2007).

Other studies on sentiment analysis have been carried out to detect sentiment from images by using a combination of images and text posted online (extracted through comments and tags) to examine how sentiment is expressed through images (Borth *et al.* 2013; Wang *et al.* 2012). Language differences between women and men have been widely discussed in Sociolinguistics. Goldschmidt and Weller (2000) combined the

[4] Some of these studies have translated text into different languages, and then adopted an English-based sentiment analysis system.

literature on gender differences in language with the literature on gender differences in emotions to shed light on the emotional talk. Their findings appear to support the stereotype that women talk more about emotions than men, whereas Grossman & Wood (1993) report that emotions are partially gender specific, with women expressing sadness more than men, and young females reporting sadness more often than males (Stapley, Haviland 1989). Aldrich and Tenenbaum's (2006) study on the verbal expression of anger, sadness, and frustration revealed that male teenagers were more likely to adopt terms linked with sadness, whilst female adolescents used words associated with frustration. These results seem to have an effect particularly on teenagers, as no significant difference was found in the verbal expression of anger, sadness, and frustration amongst middle-aged and old individuals. Bronstein *et al.* (1996) show that women are more emotionally expressive than men, and that talk more often about emotions (Goldschmidt, Weller 2000). Shimanoff (1985), however, claims that no significant gender difference surfaces when examining the verbal expression of emotions. According to Schweinberger (2019), the use of FEAR emotives is significantly more frequent in public settings, while JOY, DISGUST, and SURPRISE emotives are found to be used more in private settings.
Recent surveys on emotions have focused on gender differences in telephone speech (Cieri *et al.* 2004), e-mails (Styler 2011), letters (Mohammad, Yang 2011) to explore how males and females express emotion. Koch *et al.* (2022) investigated how the use of emoji, and emoticons, in instant messages, is stratified across gender and age. Their findings reveal that young speakers use emoticons more than old speakers. In terms of gender, emoticons were found to be mostly associated with young participants, whereas old ones use them less frequently.

Most studies on emotions, throughout the years, have been devoted to the development of a theoretical model rather than to the examination of its real-world usage, even when data was collected through word lists (Johnson-Laird, Oatley 1989). The interrelationship between language and emotion, indeed, has been mainly explored from a semantic perspective (Wierzbicka 1992). A whole line of projects on mental verbs (such as *think, wonder, imagine*, etc.) has focused on how individuals explain the distribution of the semantic roles of Cause (*e.g. That noise irritates me),* Experiencer (e.g. *That noise irritates me*), and Effect (e.g. *he trembled with fear*). Thus, so far, little quantitative research has been carried out on the language-emotion-society interrelationship.[5]

4 Methods

Joseph (2008, p. 687), in highlighting the recent developments which the discipline of Linguistics has gone through, claims that: "Linguistics has always had a numerical and mathematical side… but the use of quantitative methods, and, relatedly, formalisations and modeling, seems to be ever on the increase; rare is the paper that does not report on some statistical analysis of relevant data or offer some model of the problem at hand." Guy (1993, p. 235) states that: "The ultimate goal of any quantitative study […] is not to produce numbers (i.e. summary of statistics), but to identify and explain linguistic phenomena." Sankoff (1988b, p. 151) explains the advantage of adopting a quantitative approach in the following terms:

[5] Society, in this case study, is represented by two broad social categories: age and sex.

> […] whenever a choice can be perceived as having been made in the course of linguistic performance, and where this choice may have been influenced by factors such as nature of the grammatical context, discursive function of the utterance, topic, style, interactional context or personal sociodemographic characteristics of the speaker or other participants, then it is difficult to avoid invoking notions and methods of statistical interference, if only as a heuristic tool to attempt to grasp the interaction of the various components in a complex situation.

Along this line, the present study provides a detailed quantitative account (i.e. mixed-effects logistic regression analysis) of word-emotion lexicon by using variationist sociolinguistic methods, as discussed later in the section. Data was collected from web-based newspapers, namely *The Cook Island News* and *The Cook Islands Herald,* which contain, among others, leading articles, and editorials we decided to concentrate on. Firstly, we selected 5 female journalists and 5 male journalists born-and-bred in the Cook Islands; secondly, we randomly collected news articles written by the 10 journalists we had chosen, who were stratified by sex and age[6]. All the selected articles were written during the Covid-19 pandemic and included a variety of topics (e.g. health, politics, history, etc.).

10 Journalists				
Age			**Sex**	
Young	Middle	Old	Females	Males
3	3	4	5	5

Table 1
Sample of the present study.

In the present study, we extracted a total of 500 words, which were manually coded in an Excel spreadsheet: in one column, we coded the dependent variable (emotional words vs. non-emotional ones), whereas in the other columns we coded the independent variables such as, age, sex, word class, and word-frequency. Each emotional word was also coded according to the core emotion (anger, anticipation, fear, disgust, joy, sadness; surprise, and trust) it is associated to. Approximately, we collected 50 words per participant[7].

One of the main methodological issues is related to the coding of emotionality. Previous studies mostly adopted a polarity categorization of emotions – which appears not to be completely reliable. In pragmatic research, for instance, a more detailed coding of emotionality is usually required. The Semantic Orientation Calculator (SO-CAL) dictionary provides a more detailed scale of classification which goes beyond polarity categorization and includes additional points (see Taboada *et al.* 2011 for further details). Since terms can vary in meaning depending on the context, there is no fixed association between words and emotions. However, if numerous speakers associate a given word with a specific emotion, that word could be prototypically connected with the same emotional state. A drawback of this method is its reliance on lexical items which suggests that emotionality ratings are determined only based on the presence of a word, regardless of whether it appears in fixed-expressions or not. For instance, the lexical item “good” is regarded the same way whether it appears in a fixed expression like "*good morning*" or in statements like "*That's a very good idea*." Other drawbacks include the disregard for

[6] For recent studies with a similar number of participants see, for instance, Amos *et. al* (2020).

[7] To have sufficient data for statistical analysis, at least 30 tokens per cell are recommended (Guy 1980), as traditionally acknowledged in Sociolinguistics.

context, semantic ambiguity as well as the underlying assumption that the meaning of some lexical items is set, context-independent, and stays largely stable over time (Schweinberger 2019).

The WordNet Affect Lexicon (WAL) (Strapparava, Valitutti 2004) classifies words into six basic emotions, whereas the General Inquirer (GI) (Stone *et al.* 1966), categorizes words into various categories, including positive and negative semantic orientation.

In the present study, we use the *Word-Emotion Association Lexicon* (Mohammad, Turney 2013) for the emotion coding of sentiment analysis. The *Word-Emotion Association Lexicon* encompasses 10,170 lexical items which received a score associated to eight emotions on the basis of ratings collected through the *Amazon Mechanical Turk service* (MTurk).[8] A number of 2,216 raters were asked whether a given term was linked to one of the eight emotions (ANGER, ANTICIPATION, FEAR, DISGUST, JOY, SADNESS; SURPRISE, and TRUST) with a number of 38,726 ratings. Mohammed & Turney (2013) claim that each word was rated five times, and about 4 raters provided identical outcomes, for 85 percent of terms. Words like *cry* or *tragedy* were commonly linked with sadness, whereas terms like *beautiful* was associated with joy, and words such as *burst* signaled anger. This implies that sentiment analysis allows researchers to examine how certain core emotions are expressed rather than simply classifying whether a sentence is positive or negative. One of the advantages of using sentiment analysis is that (1) it allows the examination of emotionality together with the distinct emotions, (2) studies can be replicable since crowd-sourced term-emotion data are adopted (Schweinberger 2019).
Along the line of Schweinberger (2019), the present study applies sentiment analysis to investigate the stratification of word-emotion lexicon from a sociolinguistic perspective. Then, mixed-effects logistic regression analysis was carried out in Rbrul (Johnson 2009) – a tool which is widely recognised amongst sociolinguistic studies. Rbrul runs in R and allows us to predict the probability of emotionality based on the presence of a number of predictors "by fitting gathered data to a logit curve and modelling the effects of these multiple factors" (Tagliamonte 2006). The independent predictors, in the present paper, include age (20-30 = young; 30-50 = middle-aged; 60+ = old); sex of news writers (six males, and four females); word frequency, which is measured through the SUBTLEX-US corpus (Brysbaert, New 2009), according to which words are assigned a value which goes from 1 to 3, whereas high frequency ones are given a value of 4-7 [9]; word class (e.g. verbs, nouns, adjectives, and adverbs), words and individual news writers were also included in the statistical model as random effects. The dependent variable is emotionality, which is coded as a binary factor (emotional vs. non-emotional), and the selected application value is emotional.

5. Results and Discussion

A combination of a threefold methodology (variationist sociolinguistics, sentiment analysis, and computational methods) was adopted to find out how word-emotion lexicon is socially stratified in the Cook Islands English online news. The dependent variable of

[8] The latter is a relatively new source which is also employed in social sciences to recruit participants (Ciancia, Gallo 2021). MTurk was already adopted for linguistics and speech processing research (Yu, Lee 2014) and also recently used by D'Onofrio and Eckert (2020) to gather participants in order to investigate the icon properties of language.

[9] For a more detailed account of the corpus see Brysbaert and New (2009).

the present study is binary (emotional words vs. non-emotional ones), whereas the independent variables include age, sex, word class, and word-frequency.

Results from the multivariate regression analysis show that the most powerful predictor which reached statistical significance rejecting the null hypothesis is age, as illustrated in Table 2 below. This finding represents the model achieved in the step-up/step-down analysis. In the step-up analysis, the programme adds predictors one at a time, beginning with those which have the greatest effect on the response. This process is repeated until no more significant predictors can be added. In the step-down analysis, Rbrul fits the full model and removes those independent variables which are not significant.

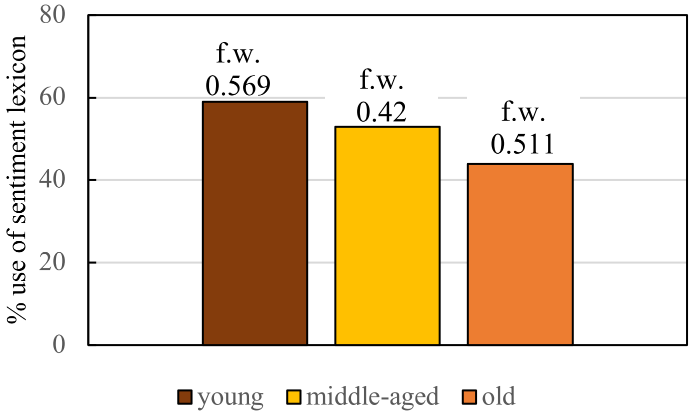
The statistical details included in Table 2 below show: R-squared (R^2), which is a measure of the goodness of fit (Winter 2020); -log-odds, which are obtained from probabilities "by taking the natural (base *e*) logarithm of the odds, where the odds are the probability of an event occurring, divided by the probability of it not occurring" (Johnson, 2009: 361);[10] and factor weights, which are the relative probabilities within the range of 0 – 1.00 and are related to the log-odds.

Application value = emotional;		**Input probability: 0.544;**			***p*-value**
Overall proportion: 0.498		**R^2 = 0.22**			
Age					***
	Logodds	**Factor weight**	**%**	***N* tokens**	
Young	0.279	0.569	59	146	
Old	0.045	0.511	44	79	
Middle-aged	-0.324	0.42	53	273	

Table 2
Mixed-effects regression analysis reporting statistically significant predictors.
Note: *p<0.5; **p<0.01; ***p<0.001.

Both Table 2 and Graph 1 show that young and old news writers are more likely to use emotional words than middle-aged ones. Thus, the use of emotional words gradually decreases from young writers to middle-aged writers until a late stage in adulthood. Note that Graph 1 measures the use of emotional words across age in terms of percentages, which differ from factor weights, as explained above. In other words, the factor weight in Graph 1 indicate that the middle-aged group remains a disfavouring factor. This finding resembles the age-grading pattern of sociolinguistic studies, according to which teenagers use non-standard features more than adults. So, the use of non-standard variants decreases from adolescence to adulthood until the old age. This significant finding ought to be replicated in a bigger sample to check for consistency.

[10] If log-odds are positive, there is a positive correlation between the variables, whereas if they are negative there is a negative correlation between them.

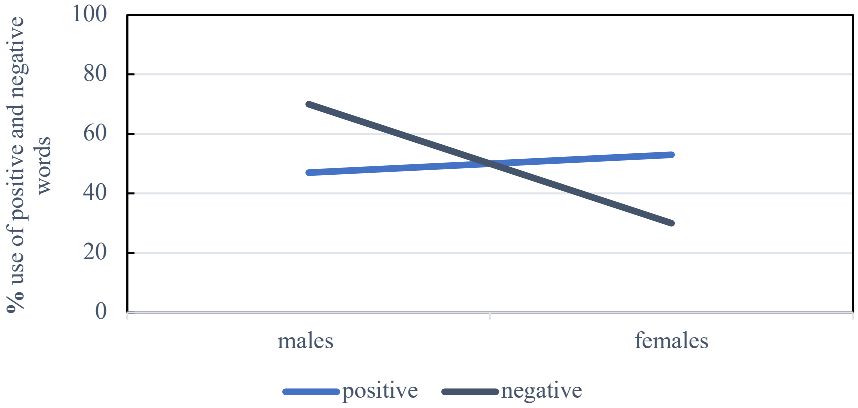


Graph 1
Use of emotional words in percentages. Factor weights are illustrated for each column.

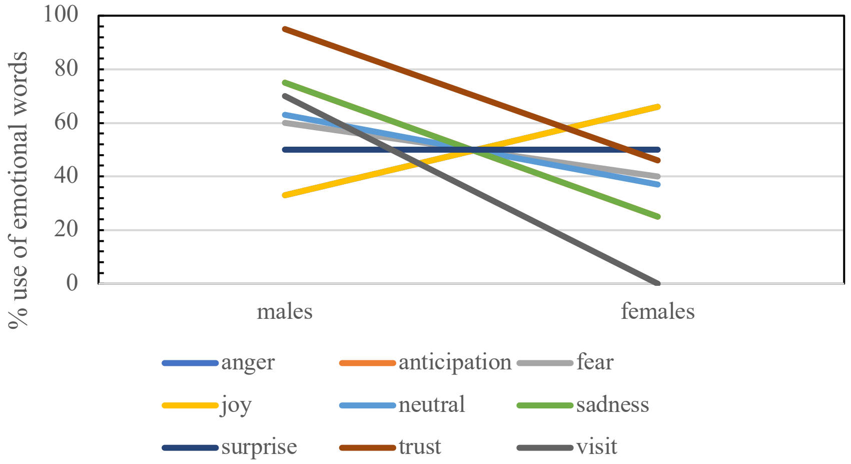
We also find that the emotions which surface the most through lexicon are *anticipation (e.g. risk, prediction,* etc.), *sadness (e.g. leave, hospital, restriction,* etc.), *fear*, (e.g. *avoid, pandemic, quarantine*) and *visit (e.g. visit)*. These findings resemble Schweinberger's (2019) results which show that *fear* words are much more likely to be adopted in public settings, whereas positive emotions occurred significantly more in private contexts. The tendency to use negative emotions, such as *fear,* does not only pertain to political discourse, but may be also indicative of a more general discursive behavior. It could also be possible that this result relates to the historical and sociopolitical context in which the data were gathered. The high use of negatively attributed emotions, such as *sadness*, and *fear*, is followed by words which index *trust* (e.g. *believe, humility, friendly,* etc.), joy (e.g. *grandchildren, congratulatory, hero,* etc.), and *surprise* (e.g. *visitors, surprise, break*, etc.). The latter also surfaces through words adopted by young news writers, with the relatively low percentage of 50% vs. 59% used by middle-aged authors. Neutral words (i.e. those which did not receive a score in the sentiment analysis by Mohammed *et al.* 2013) are used with a rate of 22% by both middle-aged and old news writers, whereas a percentage of 55 was found among young authors. Young news writers seem to adopt more words associated with *anger* than their middle-aged, and old counterparts. Words associated with anger (e.g. *opposition, loss, criticize, fighting)* tend to decrease as age increases, whereas both *trust* and *surprise* words are mostly employed by middle-aged individuals. Instead, terms that convey *joy* predominate among older authors. It seems that age as a conditioning predictor on verbal expression of emotion has been the subject of systematic empirical research only for a very few studies (Schweinberger 2019). Schweinberger (2019) discovered that older speakers use more *trust* emotives than younger speakers. It is claimed that social network for individuals over a certain age are more stable, and speakers with relatively stable networks are more likely to express their emotional state than speakers who are just beginning to form stable social bonds. Results from the present study, however, do not appear to match with Schweinberger's (2019) findings. Indeed, in more formal styles (e.g. online news) the pattern that we find, in relation to only the use of *trust* emotives, is the following:

middle-aged > young > old.

With respect to sex, Graph 2 illustrates that males seem to use more negative words than females, whilst females appear to employ slightly more positive words than men. Graph 3 suggests that the usage of *sadness, surprise, visit, anger, anticipation*, and *fear* emotions is nearly equal amongst males and females, with the exception of terms associated with *visit*, as no tokens were coded for females. Moreover, males use terms associated with *trust*, and *sadness* more than females - a finding which seems to contrast with previous studies which have shown that women tend to verbalise *sadness* more than men (Grossman & Wood 1993). In line with Schweinberger (2019), males seem to express negative emotions such as *anger*, and *fear*, whereas females tend to express more positively stereotyped emotions like *joy*.



Graph 2
Polarity categorization of emotions distributed across sex



Graph 3
Interaction between the eight core emotions and sex.

When reporting statistical results, Guy (2018) suggests that “all independent variables tested should be reported, whether significant or not. Non-significance of a potential predictor is an important finding”. In the current study, the following predictors failed to reach statistical significance: word class, gender, and word frequency. With respect to word class, despite the idea that part of speech correlates with how emotion is expressed (Schweinberger, 2019), the present study does not report any significant interaction between word class and words associated with emotions. Table 2 shows that emotional adjectives are more likely to be used in the Cook Islands English online news, and that the use of adjectives favors emotionality more than nouns, adverbs and verbs, but not significantly. Adverbs and verbs were collapsed in the mixed-effects logistic regression analysis as the number of tokens in the class of adverb did not reach the statistical threshold. Nonetheless, while prepositions appear to play a role in the expression of emotional words on a theoretical level, they were excluded not being present in the sentiment *Word-Emotion Association Lexicon* (Mohammad, Turney 2013).

When looking at sex of news writers through one level analysis, we find that females favour the use of emotional words, whilst males disfavour it, but not significantly.

Factor Group	**Logodds**	***N* Tokens**	**%**	**Factor weight**
Word class				
Adjective	0.335	62	60	0.583
Noun	-0.043	313	50	0.489
Adverb+verb	-0.292	125	43	0.428
Sex				
Females	0.138	205	60	0.583
Males	-0.138	295	50	0.489

Table 2
Non-significant predictors. Results from one level analysis.

Word frequency, included in the model as a continuum factor, did not reach statistical significance. This predictor is frequently included in linguistic studies as there is evidence that non-standard forms of language are more likely to be found amongst frequent words.

6 Conclusion

This paper has focused on the interrelationship between language, social factors, and word-emotion lexicon by investigating a small-scale corpus of 500 items, collected from 10 Cook Islanders online news writers, who were stratified by sex, and age. Emotions, as mentioned early in the paper, have been mostly examined at a theoretical level, hence little systematic research has been carried out so far. Not only has this paper provided a contribution by exploring word-emotion lexicon at the intersection with social factors in formal style, but it has presented a fine-grained coding of emotionality of the language in Cook Islands English online news.

Our findings suggest that age is a statistically significant predictor, with young and old news writers favouring the use of word-emotion lexicon in online news. The emotions which were mostly conveyed through lexicon were *anticipation (e.g. risk, prediction,* etc.), *sadness (e.g. leave, hospital, restriction,* etc.), *fear*, (e.g. *avoid, pandemic, quarantine*), and *visit (e.g. visit).* Word class, word-frequency, and sex do not seem to condition the use of sentiment lexicon in our data. When looking at the direction of sex, we found that females tend to use slightly more emotional words more than men, confirming the socially acknowledged generalisations according to which women are more likely to adopt sentiment lexicon compared to men. Our findings suggest that external social factors should be taken into account in the analysis of word-emotion lexicon. The present study will be further expanded and we will also investigate the relationship between language and emotions in the Cook Islands casual speech data by using a naturally-occuring spontaneous corpus collected on situ by Siria Guzzo in 2020.

Bionotes: Carmen Ciancia is Research Fellow at the University of Salerno. She holds a PhD in Sociolinguistics and an M.A.in English Language and Linguistics both received at the University of Essex (UK). She has presented her work at the worldwide leading conferences in Sociolinguistics (in the USA, in the UK, and in Europe). Her main research interests include sociolinguistics, language variation and change, language attitudes, endangered languages.
Siria Guzzo is Associate Professor of English Language and Linguistics at the University of Salerno, Italy. She holds a PhD in English for Special Purposes and a MA in Sociolinguistics. Her research interests mainly lie in the field of sociolinguistics and language variation and change. She has conducted research and widely published in the fields of migration and its effects on identity, new dialect/ethnolect formation, language contact and its outcomes, and first and second language acquisition. Her publications include wide-ranging investigations on the Anglo-Italian community in the UK, and a forthcoming volume on the newly-emerging Cook Island variety of English.

Author's address: cciancia@unisa.it; sguzzo@unisa.it

References

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

Aldrich N.J. and Tenenbaum H.R. 2006, *Sadness, Anger, and Frustration: Gendered Patterns in Early Adolescents' and their Parents' Emotion Talk*, in "Sex Roles: A Journal of Research" 55 [11-12], pp. 775-785.

Amos J., Kasstan J. and Johnson W. 2020, *Reconsidering the variable context: A phonological argument for (t) and (d) deletion*, in "English Today" 36 [3], pp. 6-13.

Batson C.D., Shaw L.L. and Oleson K.C. 1992, *Differentiating affect, mood, and emotion: Toward functionally based conceptual distinctions*, in Clark M.C. (ed.), *Emotion. Review of Personality and Social Psychology*, Newbury Park, CA, Sage, pp. 294-326.

Bell A. and Gibson A. 2008, *Stopping and Fronting in New Zealand Pasifika English*, in "University of Pennsylvania Working Papers in Linguistics" 14.

Benamara F., Cesarano C., Picariello A., Reforgiato D., Venkatramana R. and Subrahmanian S. 2007, *Sentiment analysis: Adjectives and adverbs are better than adjectives alone*, ICWSM, Citeseer.

Benamara F., Popescu V., Chardon B., Asher N. and Mathieu Y. 2013, *Assessing opinions in texts: Does discourse really matter?*, in Taboada M. and Trnavac R. (ed.), *Nonveridicality and Evaluation: Theoretical, Computational and Corpus Approaches*, Leiden, Brill, pp. 127-150.

Biewer C. 2015, *South Pacific Englishes: A Sociolinguistic and Morphosyntactic Profile of Fiji English, Samoan English and Cook Islands English (Varieties of English around the World G52)*, The Netherlands/Philadelphia, John Benjamins Publishing.

Biber D. and Finegan E. 1988, *Adverbial stance types in English*, in "Discourse Processes" 11, pp. 1-34.

Boucher J. and Osgood C.E. 1969, *The pollyanna hypothesis*, in "Verb. Learn. Verb. Behav" 8, pp. 1-8.

Brysbaert M. and New B. 2009, *Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English*, in "Behavior Research Methods" 41, pp. 977-990.

Ciancia C. and Gallo A. 2021, *Linguistic Fieldwork amid the Covid-19 Pandemic: The Impact of Social-distancing on Data Collection*, in "I-LanD Journal" 2, pp. 135-153.

Colombetti G. 2009, *What language does to feelings*, in "Journal of Consciousness Studies" 16 [9], pp. 4-26.

El-Beltagy SR and Ali A. 2013, *Open issues in the sentiment analysis of Arabic social media: A case study*, In "Proceedings of 9th International Conference on Innovations in Information Technology", Al Ain, UAE.

Foolen A. 2012, *The relevance of emotions for language and linguistics*, in Foolen A., Lüdtke U.M., Racine T.P. and Zlatev J. (eds), *Moving Ourselves, Moving Others. Motion and Emotion in Intersubjectivity, Consciousness and Language*, Amsterdam, John Benjamins, pp. 349-368.

Ghorbel H. 2012, *Experiments in Cross-Lingual Sentiment Analysis in Discussion Forums*, in Aberer K., Flache A., Jager W., Liu L., Tang J., and Guéret C. (ed.), *Proceedings of the 4th International Conference on Social Informatics*, Berlin, Springer, pp. 138-151.

Goldschmidt O.T. and Weller L. 2000, *Talking emotions: Gender differences in a variety of conversational contexts,* in "Symbolic Interaction" 23, pp. 117-134.

Grossman M. and Wood W. 1993, *Sex differences in intensity of emotional experience: A social role interpretation*, in "Journal of Personality and Social Psychology" 65 [5], pp. 1010-1022.

Guy G.R. and Torres Cacoullos R. 2018, *Reporting statistical results for LVC*. Paper presented at "New Ways of Analyzing Variation (NWAV)" 47, 19 October, New York.

Guzzo S. forthcoming, *Cook Islands English: a newly-emerging variety of English?*, Cambridge University Press, Cambridge.

Haas M. and Versley Y. 2015, *Subsentential sentiment on a shoestring: A crosslingual analysis of compositional classification*, in "Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics", Denver, CO.

Huang T-H., Yu H-C. and Chen H-H. 2012, *Modeling Polyanna phenomena in Chinese sentiment analysis*, in "Proceedings of COLING 2012: Demonstration papers", pp. 231-238.

Horne L. 1989, *A Natural History of Negation*, The University of Chicago Press, Chicago.

Holmes J. and Meyerhoff M. 2003, *Different Voices, Different Views: An Introduction to Current Research in Language and Gender*, in "The Handbook of Language and Gender"*,* pp. 1-17.

Johnson-Laird P.N. and Oatley K. 1989, *The language of emotions: An analysis of a semantic field,* in *"*Cognition and Emotion" 3 [2], pp. 81-123.

Kachru B.B. 1985, *Standards, codification and sociolinguistic realism: the English language in the outer circle*, in Quirk R. and Widdowson H.G. (eds.), *English in the World: teaching and Learning the*

Language and Literatures, Cambridge University Press, Cambridge, pp. 11-30.

Kachru B.B. 1992, *Teaching World Englishes*, in Kachru B.B. (ed.), *The other tongue: English across cultur*es, University of Illinois Press, Urbana/Chicago, pp. 355-365.

Kachru B.B., Kachru Y. and Nelson C.L. 2006, *The handbook of World Englishes*, Blackwell, Oxford.

Kachru Y. and Nelson C.L. 2006, *World Englishes in Asian Contexts*, University Press, Hong Kong.

Koch T.K., Romero P., Stachl C. 2022, *Age and gender in language, emoji, and emoticon usage in instant messages*, in "Computer in Human Behaviour" 126, pp. 1-12.

Krippendorf K. 2004, *Content Analysis: An Introduction to Its Methodology*, Sage, Thousand Oaks, CA.

Lindquist K.A. and Barrett L.F. 2008, *Emotional complexity*, in Lewis M., Haviland-Jones J.M. and Barrett L.F. (eds.). *Handbook of Emotions*, NewYork, Guilford, New York, pp. 513-530.

Lopez-Zafra E., Garcia-Retamero R. and Martos, M.P.B. 2012, *The relationship between transformational leadership and emotional intelligence from a gendered approach*, in "The Psychological Record" 62 [1], pp. 97-114.

López R., Tejada J. and Thelwall M. 2012, *Spanish SentiStrength as a tool for opinion mining Peruvian Facebook and Twitter*, in Setlak G., Alexandrov M. and Markov K. (ed.), *Artificial Intelligence Driven Solutions to Business and Engineering Problems*, Sofia, Ithea, pp. 82-85.

Marchand M. 2012, *État de l'art: l'influence du domaine sur la classification de l'opinion,* in "Proceedings of the joint conference JEP-TALN-RECITAL", Grenoble, France, pp, 177-90.

Mohammad S. and Yang T. 2011, *Tracking Sentiment in Mail: How Genders Differ on Emotional Axes*, in "Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)", Portland, Oregon, pp. 70-79.

Mohammad S.M. and Turney P.D. 2013, *Crowd sourcing a word-emotion association lexicon,* in "Computational Intelligence" 29 [3], pp. 436-465.

Molina-González M.D., Martínez-Cámara E., Martín-Valdivia M.T. and Perea-Ortega J.M. 2013, *Semantic orientation for polarity classification in Spanish reviews*, in "Expert Systems with Applications*"* 40, pp. 7250-7257.

Moreno-Ortiz A. and Pérez Hernández L. 2012, *Lexicon-based sentiment analysis of twitter messages in Spanish*, in "TASS, Taller de Análisis de Sentimientos en la SEPLN (Sociedad Española para el Procesamiento del Lenguaje Natural)", Castellón de la Plana, Spain.

O'Connor B., Balasubramanyan R., Routledge B.R. and Smith N.A., 2010, *From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series*, in Cohen, W.H. and Gosling S. (eds.), *Proceedings of the Fourth International Conference on Weblogs and Social Media, Washington, DC, USA, May 23-26, 2010*, The AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1536.

Omondi L. 1997, *The Language of Emotions*, Amsterdam, Benjamins

Pak A. and Paroubek P. 2010, *Twitter as a corpus for sentiment analysis and opinion mining,* in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pp. 1320-1326.

Pavlenko A. 2008, *Emotion and emotion-laden words in the bilingual lexicon*, in "Bilingualism: Language and Cognition" 11 [2], pp. 147-164.

Portner P. 2009, *Modality*, Oxford University Press, Oxford.

Salameh M. and Kiritchenko S. 2015, *Sentiment analysis after translation: A case-study on Arabic social media posts*, in *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL-2015)*, Denver, CO.

Sankoff D. 1988, *Sociolinguistcs and syntactic variation*, in Newmeyer F. (ed.), *Linguistics: The Cambridge Survey*, Cambridge University Press, Cambridge, pp. 140-161.

Schweinberger M.A. 2019, *Sociolinguistic Analysis of Emotives*, in "Corpus Pragmatics" 3, pp. 327-361.

Shimanoff S.B. 1984, *Commonly named emotions in everyday conversations*, in "Perceptual and Motor Skills" 58 [2], p. 514.

Stapley J.C. and Haviland J.M. 1989, *Beyond depression: Gender differences in normal adolescents' emotional experiences,* in "Sex Roles: A Journal of Research" 20 [5-6], pp. 295-308.

Taboada M., Brooke J., Tofiloski M., Voll K. and Stede M. 2011, *Lexicon-based methods for sentiment analysis*, in "Computational Linguistics" 37, pp. 267-307.

Tagliamonte S. 2006, *Analysing Sociolinguistic Variation*, Cambridge University Press, Cambridge.

Traugott E.C. 2010, *(Inter)subjectivity and (inter)subjectification: A reassessment*, in Davidse K., Vandelanotte L. and Cuyckens H. (ed.), *Subjectification, Intersubjectification and Grammaticalization*, De Gruyter Mouton, Berlin, pp. 29-74.

Valitutti A., Strapparava C. and Stock O. 2004, *Developing affective lexical resources*, in "PsychNology Journal" 2 [1], pp. 61-83.

Vilares D., Alonso M. and Gómez-Rodríguez C. 2015, *A syntactic approach for opinion mining on Spanish reviews*, in “Natural Language Engineering” 21 [1], pp. 139-163.

Waltinger U. 2010, *GermanPolarityClues: A Lexical Resource for German Sentiment Analysis,* in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, European Language Resources Association (ELRA), Valletta, Malta.

Wan X. 2008, *Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis*, in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, Honolulu, pp. 553-561.

Wang F., Wu Y. and Qiu L. 2012, *Exploiting discourse relations for sentiment analysis*, in *Proceedings of the 24th International Conference on Computational Linguistics (COLING), Posters* Mumbai, pp. 1311-1320.

Wang S., Jiang M., Duchesne Xavier M., Laugeson Elizabeth A., Kennedy Daniel P., Adolphs R. and Zhao Q. 2015, *Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking* in “Neuron” 8 [8], pp. 604-616.

Wiebe J., Wilson T., Bruce R., Bell M. and Martin M. 2004, *Learning subjective language* in “Computational Linguistics” 30, pp. 277-308.

Wierzbicka A. 1999, *Emotions across languages and cultures. Diversity and universals*, Cambridge University Press, Cambridge..

Winter B. 2020, *Statistics for Linguistics. An Introduction using R*, Routledge, New York.

Weller K., Bruns A., Burgess J., Merja M. and Puschmann C. 2014, *Twitter and Society* in “The Journal of Media Innovation” pp. 134-137.