Proceedings

of the 8th Conference on Computer-Mediated Communication (CMC) and Social Media Corpora (CMC-Corpora 2021)

28-29 October 2021

Radboud University, Nijmegen

Proceedings of the 8th Conference on CMC and Social Media Corpora for the Humanities (CMC-Corpora 2021)

28-29 October 2021, Radboud University (Nijmegen, the Netherlands)

Editors: Iris Hendrickx, Lieke Verheijen, and Lidwien van de Wijngaert

Published by: Radboud University
Nijmegen, 2021

Conference website: https://applejack.science.ru.nl/cmc2021/

Preface

After having to postpone the 2020 edition of the *Conference on Computer-Mediated Communication (CMC) and Social Media Corpora* because of the global COVID-19 pandemic, we are very pleased to present the proceedings of the 8th edition of the conference (CMC-corpora 2021). This conference series is dedicated to the collection, annotation, processing, analysis, and exploitation of corpora of computer-mediated communication and social media for research in the humanities and beyond. The annual event brings together language-centred research on CMC and social media in linguistics, communication sciences, media studies, and social sciences with research questions from the fields of corpus and computational linguistics, language and text technology, and machine learning.

The *8th Conference on CMC and Social Media Corpora* was hosted by Radboud University in Nijmegen, the Netherlands on 28 and 29 October 2021. Due to the aftermath of the corona pandemic, the conference was organised in hybrid format: both online and in-person participants were fully accommodated.

The conference attracted 49 submissions. Each submission was reviewed by members of the scientific committee. This committee ultimately decided to accept 22 papers and 15 posters. The contributions were presented in seven oral sessions and two poster sessions. The contributions in these proceedings cover a wide range of topics that can roughly be divided into three categories. The first category focusses on language. Both sociolinguistics and interactional linguistics are included in these proceedings. The second category of contributions take a methodological approach by focussing on text mining, corpus compilation and coding, and multimodality. The third category of contributions focusses on the application of CMC corpora research to business communication or society in general.

The program also included two invited talks: a keynote talk by Malvina Nissim (University of Groningen, the Netherlands) on the perks and perils of natural language processing and a keynote talk by Jannis Androutsopoulos (University of Hamburg, Germany) on scaling social media data in research on visual prosody. In addition, the conference was completed with a panel session in which the relation between science and practice was discussed. The conference was preceded by an online workshop on data management for creating FAIR CMC corpora, which was organised with the support of CLARIN.

This volume contains abstracts of the invited talks, short papers of oral presentations, and abstracts of posters presented at the conference. Extended papers of the best conference presentations will appear in a special issue of the journal *Psychology of Language and Communication*.

We wish to thank all colleagues who contributed to the conference and to this volume with their papers, posters, and keynote lectures. Thanks also go to the members of the International steering committee (Michael Beißwenger, University Duisburg-Essen, Germany; Steven Coats, University of Oulu, Finland; and Lisa Hilte, University of Antwerp, Belgium) for their support and help. We would also like to thank Hester Groot for her practical support. Furthermore, we wish to thank Radboud University's International Office, the Centre for Language Studies, and the Fund for Scientific Practice of Organizational Communication for their financial contribution to the conference.

While building on the results of earlier CMC-corpora conferences in Dortmund, Germany (2013 and 2014), Rennes, France (2015), Ljubljana, Slovenia (2016), Bolzano, Italy (2017), Antwerp, Belgium (2018), and Cergy-Pontoise, France (2019), we hope that the Nijmegen 2021 conference will further extend scientific knowledge on CMC and social media corpora, as well as develop the practical application of that knowledge, while underlining the relevance of the humanities perspective.

Nijmegen, October 12, 2021

On behalf of the Organising committee,
Iris Hendrickx,
Lieke Verheijen, and
Lidwien van de Wijngaert

Committees

Organising committee

Iris Hendrickx	Radboud University (the Netherlands)
Lieke Verheijen	Radboud University (the Netherlands)
Lidwien van de Wijngaert	Radboud University (the Netherlands)

International steering committee (conference series)

Michael Beißwenger	University Duisburg-Essen (Germany)
Steven Coats	University of Oulu (Finland)
Lisa Hilte	University of Antwerp (Belgium)

Scientific committee

Adrien Barbaresi	Berlin-Brandenburgische Akademie der Wissenschaften (Germany)
Steven Coats	University of Oulu (Finland)
Orphée De Clercq	Ghent University (Belgium)
Darja Fišer	University of Ljubljana (Slovenia)
Aivars Glaznieks	European Academy Bolzano (Italy)
Iris Hendrickx	Radboud University (the Netherlands)
Axel Herold	Berlin-Brandenburgische Akademie der Wissenschaften (Germany)
Lisa Hilte	University of Antwerp (Belgium)
Florian Kunneman	Vrije Universiteit Amsterdam (the Netherlands)
Nina Lauran	Radboud University (the Netherlands)
Els Lefever	Ghent University (Belgium)
Julien Longhi	Université de Cergy-Pontoise (France)
Anne Fleur van Luenen	University of Amsterdam (the Netherlands)
Harald Lüngen	Institut für Deutsche Sprache (Germany)
Maja Miličević Petrović	University of Bologna (Italy)
Nelleke Oostdijk	Radboud University (the Netherlands)
Celine Poudat	Côte d'Azur University (France)
H. Müge Satar	Newcastle University (United Kingdom)
Stefania Spina	Università per Stranieri di Perugia (Italy)
Egon W. Stemle	Eurac Research (Italy)
Erik Tjong Kim Sang	Netherlands eScience Center (the Netherlands)
Simone Ueberwasser	University of Zurich (Switzerland)
Reinhild Vandekerckhove	University of Antwerp (Belgium)
Melchior van Velzen	Radboud University (the Netherlands)
Lieke Verheijen	Radboud University (the Netherlands)
Lidwien van de Wijngaert	Radboud University (the Netherlands)

Proceedings redactor

Hester Groot	Leiden University (the Netherlands)

Table of Contents

Invited talks

The Perks and Perils of Natural Language Processing

Malvina Nissim
University of Groningen

Tools based on Natural Language Processing are by now part of daily life, both in the personal realm as well as in the work-related arena. Recommender systems on social media, automatic translators, job application assessors are just some examples. People take many such tools for granted, and oftentimes do not even realise they are playing a role, and which role, in their lives. Developing and using such tools for a variety of studies and applications poses many choices: What data should I use? Which algorithm is most appropriate? Are my results reliable? And how much are my modelling choices influencing my findings and my analyses?

In my talk, I will talk about the relationship between the giant leap forward Natural Language Processing research has made in recent years, and the risks such positive developments might pose. I will zoom in on bias and show how it can be found in models, algorithms, data, society, and - importantly - in ourselves, too. We will also see that the definition of bias is not straightforward, that what counts as bias isn't a universally accepted concept, as it heavily interacts with our own cultural and individual biases, and that it is not obvious what the best strategies are to deal with it from an NLP perspective. Biases that percolate from our society to our data and thus to our widely used language models and to the tools that are built on them are not the only problem, of course. The fact that contemporary language models, such as the GPT family, are able to generate text that can be indistinguishable from human-produced language yields potentially dangerous situations where, for example, large amounts of unaccountable machine-generated hateful comments might be spread to social media, or fake news can be beautifully concocted and massively circulated (at a volume and speed which are obviously unparalleled by human production.)

As researchers who apply computational methods to language analysis and as NLP practitioners who contribute to the development of applications used at large by society, how should we pose ourselves with respect to such issues intrinsic in, and raised by, our work? I will critically reflect on these aspects by means of examples and by interacting as much as possible with the audience.

Bio

Prof. Dr. Malvina Nissim holds a Chair in Computational Linguistics and Society at the University of Groningen, the Netherlands. She has experience in sentiment analysis and author identification and profiling, and is interested in Natural Language Generation, with a focus on style-controlled text. A crucial aspect of her work is the reflection over ethical issues in NLP, and she is a Member at Large of the newly formed ACL Ethics Committee. She is the author of 100+ publications in international venues, is member of the main associations in the field, annually reviews for the major conferences and journals, and organises and/or (co-)chairs large-scale scientific events. She graduated in Linguistics from the University of Pisa, and obtained her PhD in Linguistics from the University of Pavia. Before joining the University of Groningen, she was a tenured researcher at the University of Bologna (2006-2014), and a post-doc at the Institute for Cognitive Science and Technology in Rome (2006) and at the University of Edinburgh (2001-2005). In 2017, she was elected as the 2016 University of Groningen Lecturer of the Year.

Scaling it up on Reddit: From 'Small' to 'Bigger' Social Media Data in Research on Visual Prosody

Jannis Androutsopoulos
University of Hamburg

Research on language in social media can be broadly divided into 'small' and 'big data' approaches. Advantages and disadvantages put aside, there is a lack of theoretical and methodological dialogue between the two approaches. Research on large social media corpora draws on computational methods, while studies that rely on small data sets are becoming more hybrid, blending on and offline communication and different media formats (cf. Androutsopoulos/Staehr 2018, Bolander/Locher 2020). In this talk, I argue that 'small' and 'big' data approaches would mutually benefit from exchange and synergy-building, and suggest that one way to achieve this is in terms of research designs that start out with 'smaller' and move on to 'larger' data sets ('scaling up') or the other way around ('scaling down'). Such back and forth between aggregated data and richer situational and sequential contexts can help illuminate language in social media from several perspectives (Ilbury 2020).

This idea is discussed on the example of a case study on visual prosody, a term used here as a bracket term for manipulations of written-language signs that fulfil expressive functions (e.g. emphasis, excitement, a jocular tone), which are mainly conveyed by prosody in speech. The feature in focus is the 'agitation mark' (*Aufregezeichen*) in German social media: a sequence of graphs that consists of one or more tokens of the exclamation mark, the digit <1> and occasionally other graphs, for example <!!11!>. <!!1!11>, or <!1!!>. In an explorative study (Androutsopoulos 2022), a small and subjectively collected corpus of agitation marks was used to identify the sign's semiotic status, its pragmatic meaning, and some patterns of discourse usage. This corpus was 'subjective' – in the sense of consisting out of screenshots I made myself over several months on a range of social media platforms, including Facebook, Twitter, and Reddit – and contextually rich, as all examples came with a lot of discursive, sequential and multimodal context and were personally identified and scrutinized by myself. The qualitative analysis of this material found that agitation marks are used as a marker of stylization and distancing, and contextualize social voices – voices of populists, chauvinists, covid deniers, and others – that the writers of the sign distance themselves from. Despite considerable allographic variability, these sign combinations are recognised as tokens of a distinct graphic sign, and have developed a distinct political usage in German digital discourse in the course of the 2010s.

While 'small' data take us quite far indeed in some respects, they are not really adequate in others: a more systematic sampling of social media data is required to examine, for example, just how frequent this sign is in social media data and how its use developed over time. To this aim I am currently developing a second data collection from the biggest and oldest German-language community on Reddit, /de. My interest at this stage is to reflect on the methodological implications of this upscaling, which entails not just (obviously) a widening of the data, but also a narrowing, inasmuch as data is now collected on one particular subreddit, and a pre-configured control over sampling and selected metadata.

References

Androutsopoulos J. (in press/2022) Interpunktion und Stilisierung im digitalen Diskurs: Struktur, Registrierung und Pragmatik des ‚Aufregezeichens'. In *Sprachreflexive Praktiken – Empirische Perspektiven auf Metakommunikation*, ed. by F. Busch/ P. Droste / E. Wessels.

Androutsopoulos, J. / A. Stæhr (2018) Moving Methods Online: Researching Digital Language Practices. In *The Routledge Handbook of Language and Superdiversity*. ed. by A. Creese / A. Blackledge.

Bolander, B. / M.A. Locher (2020) Beyond the Online Offline Distinction: Entry Points to Digital Discourse. *Discourse, Context & Media* 35.

Ilbury, C. (2020) "Sassy Queens": Stylistic orthographic variation in Twitter and the enregisterment of AAVE. *Journal of Sociolinguistics* 24:2.

Bio

Prof. Dr. Jannis Androutsopoulos is professor in German and Media Linguistics at the University of Hamburg, and from 2016-2023 research professor at MultiLing, University of Oslo. His research explores relationships between language, media and society, covering themes such as spelling and script in digital communication, multilingualism online, language ideologies in media discourse, and the role of media in sociolinguistic change. Recent publications include *Digital language and literacies: practices, awareness, and pedagogy* (Guest-edited Special Issue, *Linguistics & Education*, 2021), and *Polymedia in interaction* (Guest-edited Special Issue, *Pragmatics & Society*, 2021). Together with Heike Zinsmeister, he coordinates *DiLCo*, a new research network on "Digital language variation in context".

Papers & Posters

Automatic Humor Detection on Jodel

Manuela Bergau
Radboud University Faculty of Science, Nijmegen, Netherlands
E-mail: m.bergau@student.ru.nl

Abstract

We investigated the influence of humor on the success of a post on the social media platform Jodel. Do Jodel users prefer humorous posts over other content? As posting on Jodel occurs anonymously, popularity is solely based on the content of the post. This makes it the ideal data source for our analysis. To answer our research questions we trained an SVM classifier for humor detection on hand labeled Dutch posts. We achieved a precision of 0.7 which is in line with previous work on English data. The number of user votes on the individual post was used to divide the dataset into humorous and non-humorous. Our results show that automatic humor classification is possible for Dutch text, using simple text features that were originally crafted for the English language. Furthermore, our results show that humor indeed has a high positive influence on the success of a posting.

Keywords: humor detection, classification, text feature, Dutch, virality

1. Introduction

In contrast to many other social media platforms, Jodel takes a different approach. Content posting is done anonymously, only visible geographically locally to the author, and only for a short time. For that reason Jodel calls itself a "Hyperlocal Community" (Jod, 2014). The size of the region is based on the number and activity of the users in that region. Users can either upvote or downvote a post. This respectively increases or decreases a number linked to the post which we call the vote count. Due to the anonymous nature of Jodel the decision to vote is solely based on the content of the post.

In this project we wanted to investigate the relationship between the content of a post and the vote count. Our hypothesis is that humorous posts are having a higher vote count than other posts. In order to test our hypothesis we build a humor detector based on text features.

In previous work there are a number of attempts to detect humor in an automatic way. Mihalcea and Strapparava (2005), Mihalcea and Pulman (2007), Bueno et al. (2018) and Sjöbergh and Araki (2007) investigated different text features that enable automatic humor classification.

All these papers suggest that it is possible to distinguish jokes from non-jokes with good precision using English data. The experiments vary on the scope of the used features. Some identify complex semantic patterns and some focus on simpler stylistic features e.g. the number of negations.

The data we have is mainly written in Dutch. This poses a challenge for our project while, as far as we know, only one attempt has been made before to automatically detect humor in Dutch text (Winters and Delobelle, 2020).

We investigated which features work best for automatic humor detection in Dutch language data. As such we tested a subset of features found in the literature and some we created ourselves. Furthermore we investigated the influence of different text features on the prediction.

Lastly we tried to answer the question of if Jodel users prefer humorous posts over other content.

2. Related work

There have been various attempts in automatic humor detection. The research we found have some things in common: The data is drawn from Twitter or an existing joke database and there is overlap in the used features. Besides that, the focus of most research is on English language data. Following we list the most related papers and briefly describe their focus of used features.

Mihalcea and Strapparava (2005) show that it is possible to build an automatic system that distinguishes between humorous and non-humorous one-liner input data. The authors identify stylistic features that are specific to humor. Their results show that automatic humor detection is possible, but more research into identifying more refined features is necessary. Later Mihalcea and Pulman (2007) focus on semantic features to separate humorous and non-humorous one-liners.

Similar, Sjöbergh and Araki (2007) detect humor in one-liner data but focus on simple features such as text similarity with known jokes, the ambiguity of the used words and features based on words that are often used in jokes.

In contrast to these simple features, Reyes et al. (2012) explore more high level patterns for humor and irony detection on Twitter such as unexpectedness and emotional scenarios.

Yang et al. (2015) try to identify latent structures that enable humor detection. Furthermore the authors focused on anchor detection to find key components that enable humor in a given sentence. The experimental results on both task indicate the effectiveness of their proposed latent structures on humor detection.

In the workshop paper by Bueno et al. (2018) various features described in literature are used to build an LSTM classifier for humor detection in Spanish tweets. The reported accuracy is close to the values reported in the English examples, which suggests that the used features seem to be language independent.

One other study on Dutch humor detection, Winters and Delobelle (2020), follows the approach taken in English

research (for example Mihalcea and Strapparava (2005)) and tries to distinguish jokes from non-jokes. These non-jokes are artificially generated or taken from a complete other genre such as news headlines which make them easier to detect. We on the other hand focus on a slightly more difficult task as all examples come from the same social media platform as we try distinguish real humorous posts from non-humorous posts.
Another part of our research is about people's preferences in social networks. Jodel's infrastructure is different from other social networks so we cannot directly transfer findings of Guerini et al. (2011). However, Guerini et al. (2011) argue that virality can be decomposed in several components that are independent of the infrastructure.

3. Method

To answer our question we used the following approach: The labels that are used as ground truth to train our classifier are given by human judgement. However, we only labeled part of the data.
In a second step we compared the prediction with the vote count to see if we can find a correlation between a high vote count and the humor prediction. Our goal there is to find a threshold of the vote count that predicts our division in humorous and non-humorous posts as closely as possible. We then use the vote count to divide the posts into two classes: humorous and non-humorous. Using this threshold we can label the rest of the data without human judgement. Then we use the previously trained classifier to see how this influences the performance.

3.1. Data

We performed our analysis on a data set, that was collected between 19-10-2019 and 28-1-2020. In our data capturing routine we probe the Jodel server in 5 minute intervals using the same location in Nijmegen. The server response contains roughly the first 50 posts in the 'recent' category (the number of posts varies, the reason is unknown to us.). Next to the message text, each post comes with some meta information. The meta data of interest to our research question are the current vote count, the number of replies to the post, the creation time. To this we add the timestamp when the data was retrieved from the server.

3.1.1. Preprocessing and Exploration

To get an impression of our data we perform a small data exploration and plotted a small time frame of the data to see how the number of votes develop over time (Figure 1). The x-axis shows the date and time as 'month-day hour'. The y-axis represents the vote count. It shows that the majority of posts remain with a vote count below 20, those lines in the plot appear nearly horizontal. A small fraction of posts have a different development of their vote count. The increase in vote count seems to be linear, with in some cases an intermediate plateau. If we look closely there are a few short lines of posts that have a negative vote count. Jodel posts with a vote count below -5 counts are removed from the platform. As a result those posts are only visible a short time. In general it is important to note that, except the posts that are removed, the trend of popularity is continuously increasing.

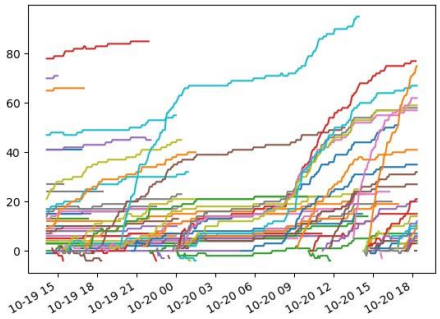


Figure 1: the development of individual vote count of posts in a small time frame

For the project we need labels about the humorous and non-humorous content. For that purpose we labeled a part of the data by hand. We ended up with four different labels: humorous, non-humorous, ambiguous, and unknown. 'Ambiguous' posts have an unclear content that might be a joke of a special non-universal kind. 'Unknown' posts are those where we are unable to identify the purpose of the author. Presumably the problem arises because the posts are written in Dutch and we are not native speakers. Later we decided to count the ambiguous and unknown label to the non-humorous class as we aim to build a binary classifier. However, especially the 'ambiguous' class could contain posts that where intended as jokes but was not recognized by the reader that labeled the data. Due to the restricted time we could not investigate this further. The test data set remains unlabeled and contains 4588 posts.
For the better readability of this paper it is important to know that we have four types of data sets: the hand labeled training, test and validation data and the unlabeled test data set. We labeled 1765 posts by hand of which we kept 765 posts for final testing purposes and the rest is used as training and validation data. We used one third of the data as validation data. The distribution of the labels across the different sets can be found in table 1.

	train set	validation set	test set
Number of samples	670	330	765
Humorous	70,9%	66,36%	73,99%
Non-humorous	17,76%	20%	19,21%
Ambiguous	6,72%	6,06%	1,7%
Unknown	4,62%	7,58%	5,1%

Table 1: Distribution of samples in the different sets

3.2. Features

The posts are represented using text features inspired by the background literature on humor detection. As all the features are crafted for English text we need to see if they work language independently. We used the following feature groups that are specific to humor detection:

- *Punctuation:* The use of punctuation, especially heavy use of exclamation and question marks are a sign for strong emotions (Bueno et al., 2018)

- *Hashtags:* Bueno et al. (2018) uses Twitter data where hashtags play a major role. In our data hashtags are present, but don't play as big a role as in Twitter.
- *Quotation:* Bueno et al. (2018) use quotations as stylistic feature. They can represent citation or a dialog.
- *Negation:* According to Mihalcea and Pulman (2007) humorous texts often contain a form of negation.
- *Adult Slang:* Mihalcea and Strapparava (2005) and Sjöbergh and Araki (2007) identify the use of adult specific words (dirty words) as a marker for humorous content.
- *Human centered:*As Sjöbergh and Araki (2007) show, many humorous posts are human centered.
- *Repetition:* Mihalcea and Strapparava (2005) use alliterations and Sjöbergh and Araki (2007) the repetition frequency of longer words as features.
- *Negative Orientation:* According to Mihalcea and Pulman (2007) a frequent use of words with a negative connotation is a marker of jokes.
- *Brackets:* We included the number of brackets as features as we noted that those are sometimes used to include or explain a joke in our data set.

4. Results and analysis

In this section we present the results of our experiments. Our analysis is structured as follows: first we present the classifiers results, trained on the hand labeled data. Then we are going to investigate the correlation of the hand-crafted labels and the vote count together with the classification results of the unlabeled data.

4.1. Classification results

We decided to use a supervised classification using SVM as the reported performance in the related work is solid. We represent the posts in a feature vector of text features as described in section 3.2.. We use the human judgement as ground truth to train the SVM classifier. The precise implementation can be found on https://github.com/ManuelaRosina/TxMM_Project.

Using all features that we described in section 3.2. we got a precision of 0.758 and an accuracy of 0.812 on the validation set. Using an ablation analysis we excluded the feature 'brackets' and 'human centered' to achieve the optimal results. We improved the precision to 0.872 with an accuracy of 0.83. All metrics are shown in table 2.

The 'brackets' feature was introduced by us and not based on background literature. The analysis showed that this did not improve the classifier's performance.

'Human centered' on the other hand was used by Sjöbergh and Araki (2007). Our assumption on why it did not work in our case is that the authors used a more clearly distinguishable data set of jokes and sentences from the British National Corpus while in our case nearly all posts are human centered. Thus our data is more nuanced than the data used in the background literature.

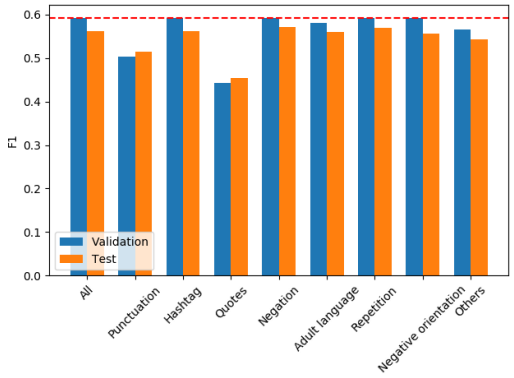


Figure 2: Influence of the different feature groups

After removing the two feature groups we run another ablation analysis, the results can be found in figure 2. The plot suggests that the groups 'Hashtag', 'Negation', 'Repetition' and 'Negative orientation' could be removed as well. However, when removing one or more of those groups the results decrease.

4.2. Handcrafted and Vote count label

In this section we present the results that are central to our main research question 'do Jodel users prefer humorous content?'. We chose the following steps in the data analysis to investigate this question: 1. Choose a threshold for the vote count based on the used metrics. 2. Compare the handcrafted- and the vote count based label: 'are they similar?'. 3. Compare the classifier metrics on both labels.

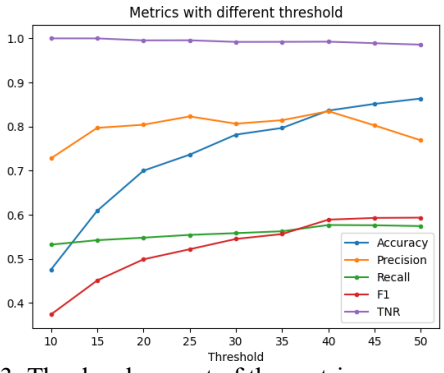


Figure 3: The development of the metrics across the range of possible thresholds

	Validation		Test		Unlabeled Test
	Hand	Vote count	Hand	Vote count	Vote count
Accuracy	0.83	0.836	0.816	0.848	**0.878**
Precision	**0.872**	0.835	0.705	0.59	0.636
Recall	**0.581**	0.577	0.559	0.535	0.545
F1	**0.593**	0.589	0.562	0.538	0.554
TNR	**0.996**	0.993	0.976	0.962	0.976

Table 2: Metrics of the test data using the two different labels

We tested the possible threshold range with the classifier which shows that the optimum is 40, as the precision drops severely after that (figure 3 shows only a part for better readability). All other metrics, except TNR, are increasing. The TNR is slightly decreasing but stays close to 1.

The results of the comparison between the handcrafted label and the vote count based label can be found in table 2. It comes as no surprise that the handcrafted labels achieve the best results across nearly all metrics, considering that the classifier was trained using them. However, in the test data the vote count label achieves a slightly higher accuracy. Also, the difference of most metrics is minimal except for the precision of the unlabeled test data, using vote count, compared to the handcrafted label. The classifiers performance on the unlabeled test data is similar to the labeled test data.

As figure 4 shows, there are a few differences in the two ways of labeling. We expected that the conflicting labels would occur mainly below the threshold. As the figure clearly shows this is not the case. For future work it would be interesting to explore the message of those conflicting labels.

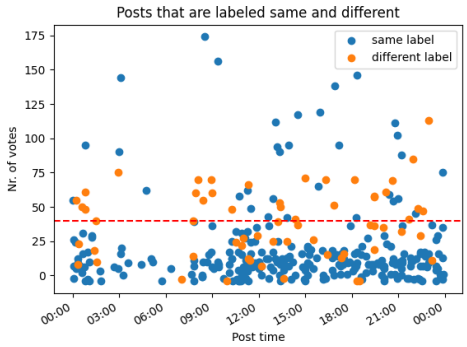


Figure 4: A comparison of the two labels using the validation data. The red line marks the vote count threshold.

5. Discussion and conclusion

Before we discuss our results we first mention some limitations of our study. Due to the lack of resources we needed to do the hand labeling ourselves, as such we cannot guarantee that no unconscious bias was introduced. While labeling we noticed that some texts are only funny in combination with previous posts which cannot be captured by our way of labeling.

Apart from these difficulties other limitations are in the nature of the data itself. As we have no information on the authors except for a vague location our results give only insights on the humor of this group of authors, but we do not know the size of this group. Based on the content and location we can make the carefully assumption that most users are students. As a premise of Jodel is the anonymity of the authors we find it not ethical to try to identify them in any way. For the same reason we did not include examples from the data.

Considering all this limitations we think that our results show, that it is possible to recognize humor in the Dutch language data with simple text features that were originally crafted for the English language. The features that were excluded during the ablation analysis were not suited for the data because of different characteristics compared to the data used in the background literature, not due to language differences. Overall the precision of around 70% is in line with the results reported by previous work.

Furthermore, our results show that there is a preference of humorous content. However, humor does not seem to be the only content characteristic that explains popularity. There is a difference between the two labels that results in a precision difference of around 10% in the test set. The outlier posts were not only below the threshold, as expected. Closer inspection of those posts makes us suspect that humor is a more nuanced category, which is not captured by a binary label. Also, popularity seems to be gained by strong opinions. However, this is pure speculation and needs to be inspected by future work.

In this paper we have shown that automatic humor classification is possible in the Dutch language using simple text features that were originally crafted for the English language. We found that there is a tendency for humorous content to gain more votes by the user than other content. As this is not the only explanation for the popularity of a post further research with a refined labeling approach is needed that takes into account the level of humor as well as strong emotional opinions (e.g. with a sentiment analysis). The influence of the timing of a post to its success could also be subject of further investigation.

6. References

Bueno, R. O., Muñiz-Cuza, C. E., Medina-Pagola, J. E., and Rosso, P. (2018). Uo upv: Deep linguistic humor detection in spanish social media. In *IberEval@SEPLN*.

Guerini, M., Strapparava, C., and Özbal, G. (2011). Exploring text virality in social networks. In Lada A. Adamic, et al., editors, *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*. The AAAI Press.

(2014). Jodel webpage. https://jodel.com. Online, accessed 30-October-2019.

Mihalcea, R. and Pulman, S. (2007). Characterizing humour: An exploration of features in humorous texts. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, pages 337–347, Berlin, Heidelberg. Springer Berlin Heidelberg.

Mihalcea, R. and Strapparava, C. (2005). Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 531–538, Vancouver, British Columbia, Canada, October. Association for Computational Linguistics.

Reyes, A., Rosso, P., and Buscaldi, D. (2012). From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74:1 – 12.

Sjöbergh, J. and Araki, K. (2007). Recognizing humor without recognizing meaning. In Francesco Masulli, et al., editors, *Applications of Fuzzy Sets Theory, 7th International Workshop on Fuzzy Logic and Applications, WILF 2007, Camogli, Italy, July 7-10, 2007, Proceedings*, volume 4578 of *Lecture Notes in Computer Science*, pages 469–476. Springer.

Winters, T. and Delobelle, P. (2020). Dutch humor detection by generating negative examples. In Proceedings of the 32st Benelux Conference on Artificial Intelligence (BNAIC 2020) and the 29th Belgian Dutch Conference on Machine Learning (Benelearn 2020).

Yang, D., Lavie, A., Dyer, C., and Hovy, E. (2015). Humor recognition and humor anchor extraction. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2367–2376, Lisbon, Portugal, September. Association for Computational Linguistics.

Online Grief Corpora: Data collection and Questions

Niklas Bodenmann, Karina Frick
University of Zurich
E-mail: niclaslinus.bodenmann@uzh.ch, karina.frick@uzh.ch

Abstract

Our project “Mourning Practices on the Internet” is part of an interdisciplinary research cluster that examines the nexus in between religion and digitalization. Our team, consisting of theologians, linguists and computational linguists, particularly investigates how people digitally express their grief on the Internet, how they verbalize expressions of condolence after the loss of a loved one or a tragic event, and how the public discourse about these grief practices is shaped and in turn shapes these practices (see "Structurations" in Giddens).

In order to investigate these questions, multiple multimodal corpora are compiled from different online sources (see e.g. Beißwenger and Beißwenger/Lüngen) where grievance practices take place, such as the very specific “digital cemeteries” like gedenkseiten.de or thematically unspecific social networks like twitter.com, where filter heuristics are applied to identify the discourse around grief.

To analyze the corpora, a circular approach with quantitative and qualitative methodologies will be used. At first, data-driven concepts (such as frequency lists, keywords or collocations) will be applied in order to detect important thematic contexts and patterns. In a second step, qualitative methods (e.g. hermeneutic and ethnographic approaches) will be used to examine and interpret the emergent patterns identified by the previous step (see e.g. Bubenhofer/Scharloth).

In connection with the compilation and the analysis of the above-mentioned corpora several questions arise, which will be addressed in the poster: How can we filter the abundance of data to obtain subsets relevant to our project while still following an inductive and data-driven paradigm? How can we make use of image data? How can we identify and classify intertextuality and citations? How can we treat characters that are used in iconic ways, such as emojis, ornamental bordures or complete ASCII artworks? These structures both interfere with standard processing but deserve scholarly attention.

Keywords: corpora, social media, grief, multimodality, corpus linguistics, practices, NLP

References

Beißwenger, M. (2018). Internetbasierte Kommunikation und Korpuslinguistik: Repräsentation basaler Interaktionsformate in TEI. In H. Lobin, R. Schneider, & A. Witt (Hrsg.), *Digitale Infrastrukturen für die germanistische Forschung* (S. 307–350). De Gruyter. https://doi.org/10.1515/9783110538663-015

Beißwenger, M., & Lüngen, H. (2020). CMC-core: A schema for the representation of CMC corpora in TEI. *Corpus*, 20. https://doi.org/10.4000/corpus.4553

Bubenhofer, N., & Scharloth, J. (2013). Korpuslinguistische Diskursanalyse: Der Nutzen empirisch-quantitativer Verfahren. In I. Warnke, U. Meinhof, & M. Reisigl (Hrsg.), *Diskurslinguistik im Spannungsfeld von Deskription und Kritik* (S. 147-168). Akademie-Verlag.

Giddens, A. (1995). *Die Konstitution der Gesellschaft. Grundzüge einer Theorie der Strukturierung* (3. Aufl.). Campus.

Register Variation in Reddit Comments – a Multidimensional Analysis

Axel Bohmann, Kyla McConnell, Hanna Mahler, Gustavo Maccori Kozma, Rafaela Tosin
Albert-Ludwigs-Universität Freiburg
E-mail: axel.bohmann@anglistik.uni-freiburg.de, kyla.mcconnel@anglistik.uni-freiburg.de,
mahlerhanna@aol.de, gustavo.maccori.kozma@saturn.uni-freiburg.de,
rafaela.tosin@anglistik.uni-freiburg.de

Abstract

Research in computer-mediated communication has recently been captivated by register differences between online platforms, often using the multi-dimensional analysis (MDA) introduced by Biber (1988), e.g. Bohmann (2020) and Biber & Egbert (2018). However, platform-internal differences have only recently received attention (e.g. Liimatta 2016, 2020 on Reddit). Building on this basis, we analyze comments published on Reddit since 2005 via the Pushshift Reddit Corpus (Baumgartner et al. 2020), extracting Biber's (1988) original linguistic features and additional, platform-specific features. We plan to use this feature annotation to implement a short-text MDA (Clarke & Grieve 2019) to find out which dimensions describe the linguistic variation found on the platform and whether the topical "subreddits" can be described as different registers. Our study will therefore add to the state of knowledge in several ways:

1. Ours is one of the first studies to regard a single comment as one text, which allows us to accurately locate linguistic variation within individual users in specific contexts of writing.
2. We train a tagger specifically to overcome previous difficulties of tagging social media data (e.g. Banga & Mehndiratta 2017), based on data from Behzad & Zeldes (2020) and Gessler et al. (2020).
3. The feature extraction script, a refined and elaborated version of Biber's (1988) initial features, written in Python, will be openly available.
4. Our long-term goal is to develop an MDA solution that captures variation within and among all (English) subreddits. In addition to providing a tool for future research on computer-mediated communication, our study adds specificity to the debate on linguistic register differences online.

Keywords: Reddit, multi-dimensional analysis, register, computer-mediated communication

References

Banga, Ritu & Pulkit Mehndiratta. 2017. Tagging Efficiency Analysis on Part of Speech Taggers. In *International Conference on Information Technology* (ICIT), 264–267.

Baumgartner, Jason, Savvas Zannettou, Brian Keegan, Megan Squire & Jeremy Blackburn. 2020. The Pushshift Reddit Dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, 14th edn.

Behzad, Shabnam & Amir Zeldes. 2020. A Cross-Genre Ensemble Approach to Robust Reddit Part of Speech Tagging. http://arxiv.org/pdf/2004.14312v1.

Biber, Douglas. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, Douglas & Jesse Egbert. 2018. *Register Variation Online*. Cambridge University Press.

Bohmann, Axel. 2020. Situating Twitter Discourse in Relation to Spoken and Written Texts: A Lectometric Analysis. *Zeitschrift für Dialektologie und Linguistik* 87(2). 250–184.

Clarke, Isobelle & Jack Grieve. 2019. Stylistic Variation on the Donald Trump Twitter Account: A Linguistic Analysis of Tweets Posted between 2009 and 2018. *PloS one*, 14(9), e0222062.

Gessler, Luke, Siyao Peng, Yang Liu, Yilun Zhu, Shabnam Behzad & Amir Zeldes. 2020. AMALGUM – A Free, Balanced, Multilayer English Web Corpus. In *Proceedings of The 12th Language Resources and Evaluation Conference*. 5267–5275.

Liimatta, Aatu. 2016. Exploring Register Variation on Reddit: A Mulit-Dimensional Study. University of Helsinki Master thesis.

Liimatta, Aatu. 2020. Using Lengthwise Scaling to Compare Feature Frequencies Across Text Lengths on Reddit. In Sofia Rüdiger & Daria Dayter (eds.), *Corpus Approaches to Social Media,* 111–130. Amsterdam, Philadelphia: John Benjamins.

The World's Emotions in Emojis

Peter Bußwolder
Independent Scholar
E-mail: busswolder@t-online.de

Abstract

The study of emojis on Twitter gained increasing attention in the last years. Little attention has been paid to the time behavior of emojis, may it be years or over the course of a single week. The purpose of this explorative study is to try to close this gap by analyzing data collected from the "Emojitracker" website over the course of two and a half years. The data was collected on a minute basis for 845 provided emojis. The emojis show a distinctive weekly behavior, with several peaks, indicating lunch period or the end of the workday. Fridays to Sundays have itself a different progress. Using the correlation matrix of the most used emojis, a hierarchical clustering has been undertaken. This indicates several unique clusters, pooling emojis with different meaning and sentiment.

Keywords: Twitter, emoji, time behavior

1. Introduction

Social media platforms such as Twitter have become a prevalent way of communication. By posting short text messages users can share their opinions, feelings or make statements about events or current topics (Li et al., 2018). This data can be a valuable source for inter alia politicians, social organizations, or researchers (Pak & Paroubek, 2010).

In the last years, Twitter users have increased their use of emojis. Emojis are pictorial representation of facial expressions, emotions, hand gestures etc. that enrich the social orientation of computer-mediated communication (Walther & D'Addario, 2001; Derks et al., 2007; Tang & Hew, 2019). They emulate the nonverbal behavior of face-to-face communication to convey social and emotional aspects of a conversation (Tantawi & Rosson, 2019).

When it comes to research focusing on the use of emojis in tweets, there has been a wide variety of topics. These comprise for example the meaning of emojis across different cultures (Barbieri et al., 2016), food-related emotions (Vidal et al., 2016), solidarity in crisis events (Santhanam et al., 2019) or political elections (Hauthal et al., 2019).

Most studies focus on single, temporarily finite events. This study tries to fill this gap by studying the most used emojis over a period of two and a half years. The source of data is the "Emojitracker" website. This site counts all emojis used on Twitter in real time (Novak et al., 2015; Rothenberg, 2013). The data was collected every minute, allowing for a detailed analysis of time-related behavior, such as weekly upswings and downturns. Also, the relation of the emojis to each other is analyzed and a cluster analysis has been conducted.

The remainder of the paper is structured as follows. In Section 2, the relevant prior literature is reviewed. Section 3 describes the data collection and pre-processing. The results are presented in Section 4, first on an aggregated basis, then individually. Section 5 summarizes this study's results and discusses limitations and potential directions for future work.

2. Related work

Twitter is a popular social media platform, allowing users to inform themselves about current events and topics and to express their thoughts and opinions publicly via short text-based messages (called tweets) (Barbieri et al., 2017). Twitter has recently attracted interest from researchers to study a series of different topics, as it offers a number of opportunities to study real-life events (e.g. elections or the stock market) (Mislove et al., 2011). Twitter users have adopted an extensive use of emojis in their tweets over the last years (Huang et al., 2008; Bai et al., 2019). Emojis are pictographs that are abstractions of facial expressions, gestures, activities, animals, plants, foods etc. These support the expression, especially of emotions or moods, in computer-mediated communications (Walther & D'Addario, 2001; Bai et al., 2019; Tang & Hew, 2019) and can be combined with plain text (Vidal et al., 2016). There is a growing number of literature focusing on emojis and its predecessor emoticons, which are non-standard orthographic features (e.g. :-) or :D) (Pavalanathan & Eisenstein, 2015). The global usage of emojis was analyzed by Ljubešić and Fišer (2016) by using geo-located tweets. Their results show that emojis are present in nearly 20% of all tweets and are used by around 38% of all users. They also ordered countries by their use of emojis and clustered them. Barbieri et al. (2016) used a vector space skip-gram model to deduce the meaning of an emoji based on the semantics in which the emoji was used. They collected more than 30 million geo-located tweets from four different countries to compare the meaning of emojis in different languages. The overall semantics seem to be the same across the involved countries, even though some emojis are interpreted differently, which may be caused by socio-demographic differences. The influence of time on the prediction and interpretation of emojis was studied by Barbieri et al. (2018) primarily in the relation to seasons. While some emojis are independent of time, others are used primarily in specific seasons, like gift, snowflake, or Christmas tree emojis in the autumn and winter or different flower and sun emojis in spring and summer. Focusing on the emotional content, Novak et al. (2015) provide a sentiment map of 751 frequently used emojis. Most emojis show a positive sentiment and the distribution is similar between the thirteen languages incorporated. The demographics of U.S. Twitter users has been analyzed by Mislove et al. (2011), showing that there is a bias to an overrepresentation of urban regions and males.

3. Dataset and Pre-Processing

To analyze the usage of emojis over the course of time, the

data was retrieved using the website "emojitracker.com" created by Matthew Rothenberg (Rothenberg, 2013). The "Emojitracker" website is counting emojis being used in all public tweets released at the social media platform Twitter. To collect the data, an application was programmed that queries the provided information every minute by connecting to the REST API made available by the website. The data was collected from June 2017 until February 2020. In this study only the data for complete months were used, comprising July 2017 to January 2020. 845 emojis are covered in total. Please mind that this does not cover all emojis that are available to users on Twitter. Currently (with the latest update at the end of January 2020) users can choose between 3.243 emojis when writing a tweet (Emojipedia, 2020).

To pre-process the data, a three-step approach was chosen. First, as the website delivers the entire number of used instances for each emoji, the difference between one minute and the next is calculated. The change for every minute is the primary parameter on which the analysis is based. Second, due to technical issues, the data for a several days could not be obtained. Additionally, around 2.750 data points are noticeable low or high. This may be due to problems on the supply side (i.e. technical issues). These data points were excluded from the analysis. For the entire time of the period under review, around 2.9% of the possible data points got lost. The third and final step was the aggregation of the data. As the evaluation based on minutes is computationally intensive and is difficult to comprehend in a long-term investigation, the mean of the minute change for every hour of each day were computed. The final data set consists out of 21.936 time-related data points for each of the 845 emojis.

Beyond the data correction one emoji was excluded due to its behavior, indicating automated Twitter accounts. This is the recycling symbol (♻). The emoji shows one of the highest usages of all considered emojis until the beginning of June 2018. After this, the usage drops from 1.101 to 124 emojis per minute in average. From August 2019 it drops again to an average level of 4. Users can subscribe to apps, to send automated messages through their profile. In the Arabic world this is commonly done to keep the profile of a deceased person active. The use by muslims is supported by the fact that Friday is the day with the highest number of counted recycling symbols and the green color of it. Green is of importance in Islam (Rothenberg, 2017). On Friday the Friday prayers happen, an exalted Islamic ritual. The services have been deactivated by the users or by Twitter itself (Roth & Harvey, 2018).

To further focus the analysis, only the 25 most used emojis are considered (see Table 1 for an overview). These 25 emojis cover 50,75% of all used emojis in the designated period of time, indicating a right-skewed distribution.

The time is calibrated to Eastern Standard Time (EST), as the USA is the single biggest market regarding user base and advertising revenue (Twitter Inc., 2020).

4. Analysis and Results

The analysis is conducted on two levels. First, the time behavior of all emojis together is considered; second, the mostly used emojis are examined on an individual level.

4.1 Cumulated Time Behavior

The progress of the emoji usage on Twitter over the course of an average week by hour is shown in Figure 1. The time is set to East Coast Time. The different weekdays have significant similarities in their typical structure. The lowest point happens around 4 a.m., while the highest point is always at 12 o'clock. The mean value for both does not differ much between the days. A second peak is visible around 5 p.m. on all days except for Friday to Sunday with only marginal peaks at this time. Two explanations are likely, which cause the behavior either alone or in conjunction. The first explanation may be that the west coast of the U.S. is increasingly twittering due to the time shift. The second explanation would be that this peak displays the end of the working day, which offers time for users to twitter about current events.

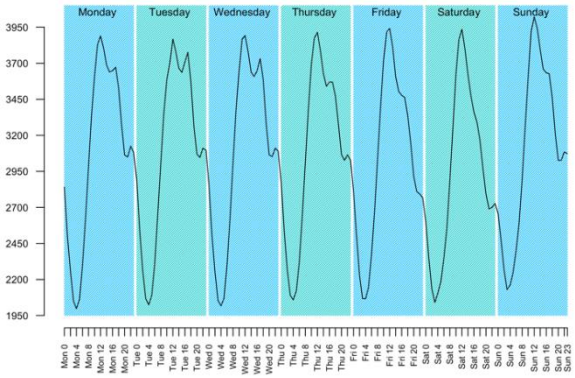


Figure 1: Average usage of emojis per hour and weekday

The second argument may explain why Friday to Sunday do not show this peak, as this distinct break is not present (e.g. due to an earlier end of the workday on Fridays) or purely free time (for Saturdays and Sundays).

The last peak, also in absolute terms already reduced, happens at 10 p.m. It is visible throughout the week before the mean usage per minute drops sharply until the beginning of the next day. Focused on the U.S. time zone, this may be caused by evening activities and could indicate for the individual user the conclusion of the day. The lower number of used emojis in total on Friday and Saturday and the less pronounced peak could be the result of weekend activities with family or friends (i.e. visiting a restaurant, going to the cinema, etc.) that outdo tweeting.

4.2 Most used Emojis and Cluster

Taking a closer look at the individual emojis in Table 1, it becomes apparent that most emojis are either facial expression or heart shapes. The first symbol is fire and is the thirteens emoji in the list and one of only 5 emojis not part of the mentioned two categories.

Pictogram	Short name	Mean usage per minute	SD
😂	face with tears of joy	707,3	175,6
❤️	red heart	370,8	104,3
😭	loudly crying face	248,1	82,3
😍	smiling face with heart-eyes	233,4	59,9
😊	smiling face with smiling eyes	122,0	30,3
💕	two hearts	92,0	26,1
😉	winking face	81,3	26,6
😘	face blowing a kiss	80,9	20,9
😅	grinning face with sweat	80,7	23,8
😔	pensive face	80,5	35,7
😁	beaming face with smiling eyes	79,1	25,9
♥️	heart suit	73,6	28,6
🔥	fire	73,5	89,3
👍	thumbs up	71,6	27,9
👀	eyes	70,7	28,6
😩	weary face	70,2	20,9
💔	broken heart	69,2	26,8
💙	blue heart	63,8	25,3
💜	purple heart	60,6	40,1
🙏	folded hands	58,9	22,7
😎	smiling face with sunglasses	58,6	19,8
😢	crying face	58,3	21,1
😏	smirking face	50,3	16,6
😒	unamused face	49,0	16,6
✨	sparkles	48,8	26,8

Table 1: 25 most used emojis with descriptive statistics

Also, the list is predominantly characterized by emojis with a positive sentiment. Only five emojis show a negative sentiment as determined by Novak et al. (2015). These are loudly crying face 😭, pensive face 😔, weary face 😩, broken heart 💔 and unamused face 😒. The eyes 👀 and crying face 😢 have a nearly neutral evaluation, even though the latter shows a bigger spread from positive to negative meaning.

Based on the minute-by-minute data, the correlation matrix is calculated (Figure 2). All correlations are positive and highly significant (p <.001).

Attracting attention are the three emojis with comprehensively lower correlations to the other emojis. There is first the fire emoji (🔥) with the lowest average, followed by the sparkles (✨) and the purple heart (💜). All three emojis show a less predictable behavior, significantly increasing the usage for a short period of time. This is also evident by the increased coefficient of variation (referring to Table 1).

By means of a hierarchical clustering, the emojis got grouped into seven cluster. The three aforementioned emojis form each a cluster of their own. The fourth cluster consists of 💔 and 😢, with a negative or neutral sentiment, respectively. This cluster may relate to breakups. The next cluster combines two negative emojis (😩 & 😒) with emojis related to love (💕, ❤️ & 😍). This may relate to love relationships, but this is difficult to substantiate without context.

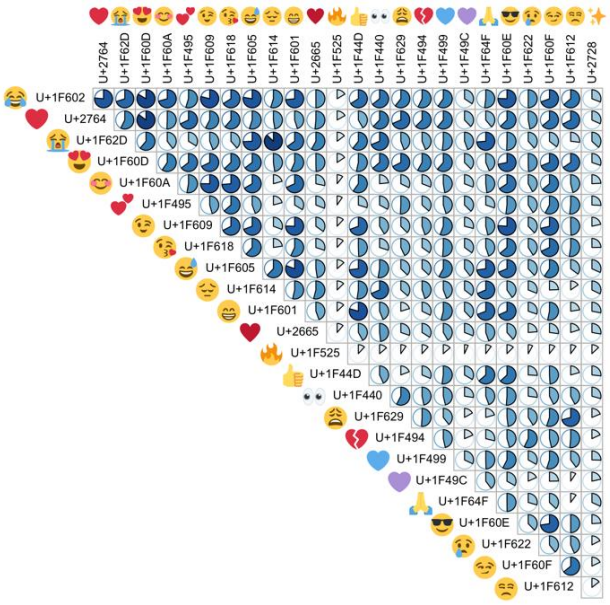


Figure 2: Correlation matrix

The sixth cluster again combines two negative emojis (😭 & 😔), now with three symbols (👀, ❤️ & 🙏). Here, sadness and sympathy may be indicated. The remaining emojis form the biggest cluster, consisting out of ten emojis, in which eight are positive facial expressions (😊, 😘, 😏, 😎, 😂, 😉, 😅 & 😁). The other two are a heart shape (💙) and a hand gesture (👍). This cluster seems to be referring to friendly relationships, inter alia with positive support. The clusters are comparable to previous results (Ljubešić & Fišer, 2016; Barbieri et al., 2016), though, due to the restricted number of emojis considered in this study, limited mostly to facial expression.

5. Conclusion

The interest in Twitter as a research subject has increased significantly over the past years. The opportunities are manifold, especially in social and linguistic research areas. However, most existing work related to emojis does not include the temporal relation of the usage of emojis. This paper tried to shed further light into this topic. The usage of emojis differs over time and is not strictly related to the number of users, but also to their activity. Moreover, the users restrict themselves to a lower number of unique emojis, that account for a disproportional part of all used emojis. Most of the here considered emojis are facial expression and have a positive meaning. This is consistent with previous research (Cappallo et al., 2018). The cluster analysis revealed several emojis with a considerable different and temporal limited behavior.

Two limitations need to be discussed that restrict the meaningfulness of this study. First, the data does only include the usage figures and no further information, like the context in which an emoji was used or in what combination it may be used with other emojis. Second, any locational information is absent. No attribution to a country or time zone is possible.

Going forward, this study sets the foundation for future work focusing on the temporal usage of emojis. This includes the broadening of the considered emojis and their relation to real world events (planned ones, like international sport events or unplanned ones, like catastrophes). Building on this, the sentiments are also of relevance and how people express their emotions collectively and over time on Twitter. This encompasses single events but also entire years (Golder & Macy, 2011).

6. Acknowledgment

The author likes to thank Stefan Becking for developing the program to collect the data from the Emojitracker REST API.

7. References

Bai, Q., Dan, Q., Mu, Z., Yang, M. (2019). A Systematic Review of Emoji: Current Research and Future Perspectives. Frontiers in Psychology, 10, pp. 2221.

Barbieri, F., Ballesteros, M., Saggion, H. (2017). Are Emojis Predictable? arXiv: 1702.07285 [cs.CL]. Retrieved from https://arxiv.org/abs/1702.07285

Barbieri, F., Kruszewski, G., Ronzano, F., Saggion, H. (2016). How Cosmopolitan Are Emojis? Exploring Emojis Usage and Meaning over Different Languages with Distributional Semantics. Proceedings of the 24th ACM international conference on Multimedia, pp. 531-535.

Barbieri, F., Marujo, L., Karuturi, P., Brendel, W., Saggion, H. (2018). Exploring emoji usage and prediction through a temporal variation lens. arXiv preprint arXiv:1805.00731.

Cappallo, S., Svetlichnaya, S., Garrigues, P., Mensink, T., Snoek, C.G.M. (2018). The New Modality: Emoji Challenges in Prediction, Anticipation, and Retrieval. IEEE Transactions on Multimedia, 21(2), pp. 402-415.

Derks, D., Bos, A., von Grumbkow, J. (2007). Emoticons and social interaction on the Internet: the importance of social context. Computers in Human Behavior, 23(1), pp. 842-849.

Emojipedia (2020). Twitter Twemoji 12.1.5. https://emojipedia.org/twitter/twemoji-12.1.5/ Accessed March 2020.

Golder, S., Macy, M. (2011). Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. *Science*, 333(6051), pp. 1878-1881.

Hauthal, E., Burghardt, D., Dunkel, A. (2019). Analyzing and Visualizing Emotional Reactions Expressed by Emojis in Location-Based Social Media. *ISPRS International Journal of Geo-Information*, 8(3), p. 113.

Huang, A., Yen, D., Zhang, X. (2008). Exploring the potential effects of emoticons. *Information & Management*, 45(7), pp. 466-473.

Li, M., Ch'ng, E., Chong, A., See, S. (2018). Multi-class Twitter sentiment classification with emojis. *Industrial Management & Data Systems*, 118(9), pp. 1804-1820.

Ljubešić, N., Fišer, D. (2016). A Global Analysis of Emoji Usage. *Proceedings of the 10th Web as Corpus Workshop*, pp. 82-89.

Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., Rosenquist, J.N. (2011). Understanding the Demographics of Twitter Users. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*.

Novak, P., Smailović, J., Sluban, B., Mozetič, I. (2015). Sentiment of Emojis. *PLoS ONE*, 10(12), e0144296.

Pak, A., Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *LREc*, 10(2010), pp. 1320-1326.

Pavalanathan, U., Eisenstein, J. (2015). Emoticons vs. Emojis on Twitter: A Causal Inference Approach. arXiv: 1510.08480 [cs.CL]. Retrieved from https://arxiv.org/abs/1510.08480

Roth, Y., Harvey, D. (2018). How Twitter is fighting spam and malicious automation. https://blog.twitter.com/official/en_us/topics/company/2018/how-twitter-is-fighting-spam-and-malicious-automation.html Accessed April 2020.

Rothenberg, M. (2013, December 9). How I Built Emojitracker. *Medium.com*. https://medium.com/@mroth/how-i-built-emojitracker-179cfd8238ac Accessed August 2021.

Rothenberg, M. (2017, November 15). Why the Emoji Recycling Symbol is Taking Over Twitter. *Medium.com*. https://medium.com/@mroth/why-the-emoji-recycling-symbol-is-taking-over-twitter-65ad4b18b04b Accessed August 2021.

Santhanam, S., Srinivasan, V., Glass, S., Shaikh, S. (2019). I Stand With You: Using Emojis to Study Solidarity in Crisis Events. arXiv: 1907.08326 [cs.SI]. Retrieved from https://arxiv.org/abs/1907.08326

Tang, Y., Hew, K.F. (2019). Emoticon, Emoji, and Sticker Use in Computer-Mediated Communication: A Review of Theories and Research Findings. *International Journal of Communication*, 13, pp. 2457-2483.

Tantawi, Y., Rosson, M. (2019). The Paralinguistic Function of Emojis in Twitter Communication. *Social Media Corpora for the Humanities (CMC-Corpora2019)*, pp. 68-72.

Twitter Inc. (2020). Selected Company Metrics and Financials. Retrieved from https://investor.twitterinc.com/financial-information/quarterly-results/default.aspx

Vidal, L., Ares, G., Jaeger, S. (2016). Use of emoticon and emoji in tweets for food-related emotional expression. *Food Quality and Preference*, 49, p. 119-128.

Walther, J., D'Addario, K. (2001). The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication. *Social Science Computer Review*, 19(3), pp. 324-347.

SUNCODAC. A Corpus of Students' Forum Discussions in Higher Education

Mario Cal-Varela, Francisco Javier Fernández-Polo
University of Santiago de Compostela
E-mail: mario.cal@usc.es, xabier.fernandez@usc.es

Abstract

Participating in asynchronous online discussions is an important ability for university students, often in increasingly multicultural environments and in a lingua franca, which imposes extra constraints that are worth exploring through samples of authentic materials (Swan & Shea, 2005; Wise et al., 2012). SUNCODAC intends to open a new window into this increasingly important reality.

SUNCODAC contains the online discussions produced in an undergraduate translation course in a Spanish university. Participants include both local and exchange students. Discussions were regularly organized through the virtual learning environment (Moodle) around a translation uploaded by a volunteer student, with classmates discussing pros and cons and suggesting improvements. The corpus includes all textual material produced in the events, including lecturers' instructions and final reflections.

SUNCODAC contains 61 full forum discussions in four consecutive years (two in Spanish, and two in English) totalling just under 600,000 words in 3,305 messages, produced by 520 students and 2 instructors.

Messages were initially stored in a database, which resulted in the loss of some font style features. No grammatical or spelling errors were edited. Metadata on users (sex, L1) and post context (date, time, language, etc.) were collected to be used for search purposes as filter variables. The materials were subsequently stored in XML to facilitate consultation and future annotation. Delimiting tags were used to identify different threads, individual messages and sections within messages (e.g. openings and closings). Texts were fully anonymized and personal names replaced by user codes.

The query tool allows the user to explore word frequencies and combination patterns. Full texts can also be retrieved according to specified criteria, making it possible to reconstruct entire threads. User-friendly dropdown menus are used for filter variables: student's sex and mother tongue, message type (initial draft, peer feedback, etc.), date and main language of the post, etc.

The materials in SUNCODAC make it possible to explore a wide range of aspects, including lexico-grammatical features of online interaction, recurring structural patterns of discussion posts, interpersonal features, particularly so-called "social presence markers" (Rourke et al., 1999), or politeness and attenuation strategies, among others. Translation studies scholars may also find it interesting to look at the content of the discussions, focusing, for example, on translation errors, references to sources of evidence and authority to support translation decisions, et cetera.

Keywords: corpus, online discussions, higher education

References

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing Social Presence In Asynchronous Text-based Computer Conferencing. *Journal of Distance Education*, *14*(2), 51–70.

Swan, K., & Shea, P. (2005). The development of virtual learning communities. In R. S. Hiltz & R. Goldman (Eds.), *Learning together online: Research on asynchronous learning networks* (pp. 239–260). Lawrence Erlbaum Associates.

Wise, A. F., Perera, N., Hsiao, Y.-T., Speer, J., & Marbouti, F. (2012). Microanalytic case studies of individual participation patterns in an asynchronous online discussion in an undergraduate blended course. *Internet and Higher Education*, *15*, 108–117.

Metapragmatic Discourses in Instagram Accounts on *Heterocringe*

Alice Cesbron
Universität Greifswald
E-mail: alice.cesbron@hotmail.fr

Abstract

Similarly to the practice of sharing "funny" memes or tweets, the practice of sharing *heterocringe* content has been growing on several social media platforms since 2019 and consists of sharing content displaying cringeworthy heterosexuality. As part of my ongoing PhD project in Queer Linguistics (Motschenbacher & Stegu, 2013) and Digital Humanities, I here want to analyze the signs and elements of metapragmatic discourse within Instagram discourses on *heterocringe*.

This analysis is based on a multimodal and multilingual corpus from Instagram, which consists of 2807 posts mostly in English and German. Using Python, these posts were collected up until March 2021 from two accounts focused on sharing *heterocringe* content. Combining Multimodal Critical Discourse Analysis (Machin & Mayr, 2012) with the tools of Corpus Linguistics, this part of my research project adopts a quantitative and qualitative methodology. The quantitative analysis focuses on word frequency, n-grams, key words and key n-grams, as well as collocations. The qualitative analysis includes Critical Discourse Analysis (Fairclough, 2014) focused on ideologies and an analysis of memes (Yus, 2019), pictures, videos and digital symbols.

In this abstract I focus on quantitative preliminary results based on the posts' captions. Contrarily to my expectations, the results were not categorically dominated by words on gender and sexuality. The n-gram and key n-gram frequency lists revealed an important number of what seem to be metapragmatic elements such as verbs describing mental processes like "I mean", "I think" and "u know" and phrases like "post this" or "this page". These elements could be interpreted as expressing the creator's opinion towards the content, highlighting the discursive situation with the followers and could also be signs of a reflection on the account. These preliminary results could thus be a starting point for a deeper analysis on the metapragmatic nature of discourses on *heterocringe* content.

Keywords: discourse analysis, gender and language, social media, corpus linguistics, metapragmatic discourse

References

Fairclough, N. (2014). *Language and Power* (3rd ed.). Routledge.

Machin, D., & Mayr, A. (2012). *How to Do Critical Discourse Analysis : A Multimodal Introduction*. SAGE Publications Ltd.

Motschenbacher, H., & Stegu, M. (2013). Queer Linguistic approaches to discourse. *Discourse & Society*, *24*(5), 519-535. https://doi.org/10.1177/0957926513486069

Yus, F. (2019). Multimodality in Memes : A Cyberpragmatic approach. In P. Bou-Franch & P. Garcés-Conejos Blitvich (Eds.), *Analyzing Digital Discourse : New Insights and Future Directions* (105-132). Palgrave Macmillan.

A Database of North American Multiple Modals from YouTube

Steven Coats
University of Oulu
E-mail: steven.coats@oulu.fi

Abstract

The use of multiple modal auxiliary verbs in a single verbal phrase is a feature of dialectal spoken English that has attracted substantial research attention, but due to its rareness, most studies of its syntax and geographical distribution have relied on non-naturalistic data with a limited geographical scope. This paper presents a database of sequences of two modal verbs that have been manually annotated as either authentic double modals (i.e. two modal verbs in a single verbal phrase) or as instances of correction/self-repair in which a speaker substitutes one modal form for another within a single speech turn (https://stcoats.github.io/dm_table1.html). The data, from the Corpus of North American Spoken English (CoNASE, Coats 2021), is linked to the corresponding video at the time of utterance. The resource attests new locations and new combinations of double modals and may prove to be useful for research into the syntax and geographical distribution of an emblematic North American English dialect feature.

In addition, the database may be suitable for corpus-based approaches to pragmatics, for example by providing data for the study of corrections or self-repair phenomena in speech. Self-repairs have been seen in the field of Conversation Analysis as a device for speaker turn regularity maintenance or local meaning negotiation (Fox et al., 1996; Schegloff et al., 1977), but have mainly been researched on the basis of small numbers of conversations. In this data, self-repairs could be considered from a quantitative perspective in terms of (for example) geographic distributions or the discourse content in which they occur. In addition, the methods used to create the database can be implemented for the creation of time-aligned multimodal corpora from publicly accessible social media videos suitable for the investigation of a wide range of syntactic, lexical, pragmatic, and interactional phenomena.

Keywords: dialectology, modal verbs, corpus linguistics, repair, conversation analysis

References

Coats, S. (2021). Corpus of North American Spoken English (CoNASE). http://cc.oulu.fi/~scoats/CoNASE.html

Fox, B. A., Hayashi, M. & Jasperson R. (1996). Resources and repair: A cross-linguistic study of syntax and repair. In E. Ochs, E. A. Schegloff & S. A. Thompson (Eds.), *Interaction and Grammar*, pp. 185–237. Cambridge: Cambridge University Press.

Schegloff, E. A., Jefferson, G. & Sacks H. (1977). The preference for self-correction in the organization of repair in conversation. *Language* 53, pp. 361–382.

Controlling Social Media Data: a Case Study of the Effect of Social Presence on Consumers' Engagement with Brand-generated Instagram Posts

Anna Corone, Annemarie J. Nanne, Emiel van Miltenburg
Tilburg University
E-mail: anna.corone@gmail.com, a.j.nanne@tilburguniversity.edu, c.w.j.vanmiltenburg@tilburguniversity.edu

Abstract

Research in social media marketing studies ways to increase customers' engagement with brand-generated social media posts. This can either be done through experiments, or corpus studies of existing social media posts. Experiments have the advantage that they are controlled, but they often lack ecological validity, while for corpus studies the reverse is often true. As a case study, we construct a corpus of 1761 brand-generated Instagram posts, looking at the effect of social presence (the perception of human contact) on different engagement metrics (likes and comments), taking the effect of possible confounds (theme of slogans, funniness, time) into account. We show how social media posts can be analyzed at different levels of granularity, to establish the strength of the effect of social presence. We hope that our work will help others to isolate the impact of different variables on post engagement on social media.

Keywords: social presence, customers' engagement, brand-generated social media posts

1. Introduction

Social presence (Short, Williams & Christie, 1976) is perceived when computer-mediated communication (CMC) tools are able to prompt the feeling of human contact through features enhancing warmth, personalization and sociability (Yoo & Alavi, 2001). Multiple studies focusing on social presence have shown that it is one of the main factors enhancing customers' engagement with social media posts (i.e., increasing the number of likes and comments) (Bakhshi, Shamma & Gilbert, 2014; Cyr et al., 2009). This is important because engagement is a sign of customers' satisfaction and emotional involvement with the brand, and therefore, it is the key to successful marketing strategies (Pansari & Kumar, 2017).

1.1 Related work

The presence of a face is an important feature that can influence social presence. The high number of neurons involved in the processing of a face resolves in increased attention to a stimulus when it contains a face compared to when it does not (Droulers & Adil, 2015). Therefore, the fact that faces are rich stimuli calls for a differentiation between levels of social presence intended as human cues in general and human faces in particular.

Earlier studies on social presence in computer-mediated-communication fall into two categories: controlled experiments and large-scale corpus studies. Cyr et al. (2009) provide an example of a controlled experiment. They distinguish three levels of social presence (low, medium, and high), and present users with carefully manipulated stimuli, to see if social presence affects user trust in e-commerce websites. An important contribution of their study is that while high levels of social presence (display of human faces) were the most appreciated, participants found medium levels of social presence (display of parts of the human body other than the face) confusing, but still better than the complete absence of a person. The downside of this study is that it is unclear to what extent their results generalize to real-world social media behavior. A way to examine real-world social media behavior is by a large-scale corpus study such has the one from Bakhshi, Shamma and Gilbert (2014). They automatically detect faces in a corpus of 1 million Instagram posts, and show that the presence or absence of faces is correlated with the number of likes and comments. While this study does provide an analysis of real-world social media behavior, because of the scale of their work, the data is relatively uncontrolled, and the authors are not able to use Cyr et al.'s more fine-grained distinction of low, medium, and high social presence. This paper explores the middle ground between the two approaches, where we select a specific set of posts, and carry out a controlled analysis of the individual factors influencing user engagement.

Figure 1: Overview of the Shanty Biscuits Instagram feed (@shantybiscuits), which is used for our case study.

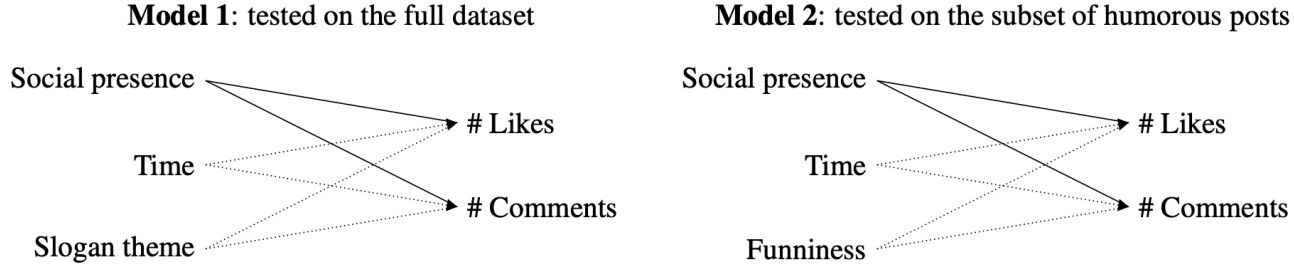


Figure 2: The two models used in our study. Both models use *social presence* as an independent variable, and *the number of likes/comments* on the post as a dependent variable. The models differ in their covariates (indicated with dotted lines). For Model 1, the covariates are *time since the post was placed*, and the *theme of the slogan*. For Model 2, covariates are *time since the post was placed*, and the *funniness of the post* (based on 10 ratings per post, see §2.2).

We present a case study, looking at the Instagram feed of Shanty Biscuits (@shantybiscuits), an existing French brand of personalized biscuits. Customers may submit slogans no longer than 12 characters including spacing or a logo to be printed on the biscuits. The company's Instagram feed is full of pictures of biscuits, apart from the text these posts differ relatively little from each other. There may or may not be a person holding the biscuit, and the backgrounds may be slightly different, but otherwise all the posts are very similar, as displayed in Figure 1. This allows for us to carry out a controlled study of the factors influencing user engagement, where our primary interest is the impact of low (no person visible), medium (part of a person visible), or high (face visible) social presence. Our study controls for the themes of the slogans, the funniness of the slogans, and the time when the picture was posted on Instagram. Based on previous studies, our **hypothesis** is that images with high social presence lead to more engagement (likes/comments) than images with low social presence and medium social presence (1a), which in turn leads to more engagement than pictures with low social presence (1b).

Thanks to the controlled setting of our study, we were able to isolate the effect of social presence from other confounds. This enabled us to examine real-world social media behavior while having a fine grained and controlled approach that helped us deal with the possible noise in the data. We found that the theme of the slogans can significantly affect whether social presence influences customers' engagement. This means that when analyzing the effect of social presence, it is important to consider also other potential factors that may confound its effect. We hope that our work contributes both to the study of social presence, and that our approach will help others aiming to isolate the impact of different variables on post engagement on social media.

2. Methods

With its people-centered focus, Instagram favors a higher degree of social presence (Bakhshi et al., 2014), making that social media platform suitable for the analysis here proposed. We carried out a content analysis of the entire Instagram feed of Shanty Biscuits. The metadata of the posts was retrieved through InstaLooter[1] on November 6, 2019. This data included the caption, the number of likes and comments, the date of publication, the Unix timestamp (the number of seconds elapsed since January 1, 1970 on the basis of which we calculated the number of days since the post was posted), the image URLs, the post URLs and the post ID of each published post. Subsequently, in order to delete possible duplicates among the posts, we used the Python Image Library[2] to find out which pictures had the exact same 200 pixels in a row, at a height of 200 pixels into the image. The present research proposes the analysis of the same hypothesis at two levels of granularity. For this purpose, two models as shown in Figure 2 are proposed. **Model 1** is tested on the whole dataset of the feed of the brand and includes social presence as independent variable, number of likes and comments as dependent variables, and the theme of the slogans of the cookies and time as covariates. **Model 2** is tested on the subset of humorous posts of the feed and includes social presence as independent variable, number of likes and comments as dependent variables and funniness perception of the posts and time as covariates.

2.1 Corpus study

The content analysis of a corpus of 2010 Instagram posts of Shanty Biscuits published between October 10, 2013 and September 30, 2019 was performed. After excluding 12 double posts with the Python Image Library, the remaining posts were manually coded in order to categorize them in three social presence categories and six text theme categories that were established throughout the analysis. A single annotator coded the Instagram posts. After the first round of coding, the same coder recoded the posts again in order to ensure reliability. At the end of the coding process, 231 posts were excluded as their pictures did not display the product and six posts were excluded as they were slideshows. At this stage, the final corpus consisted of 1761 brand-generated Instagram posts.

2.1.1. Social Presence

The three levels of social presence were coded consistent

[1] See: https://instalooter.readthedocs.io/en/latest/

[2] https://pillow.readthedocs.io/en/stable/

with the research of Cyr and colleagues (2009). The content analysis lead to the following categorization. **High social presence level** (*n*=146): included all the posts where the product appeared in the foreground and human facial features appeared entirely, partly or blurred in the background. **Medium social presence level** (*n*= 379): included posts in which parts of the human body other than the face were displayed. Most of the posts of this category show the product while being held by a hand. **Low social presence level** (*n*=1236): included posts where only the product appeared in a neutral background and no human images where displayed.

2.1.2. Theme of the slogans

With regards to the theme of the slogans, the analysis of the posts led to six categories. **Humorous text** (*n*=271): the slogans on the cookies of this category included funny quotes from movies, puns, sarcastic, silly, or mocking statements, and references to everyday problems. **Emotional events of life** (*n*=211): the slogans of the cookies of this category referred to the celebration of important events such as weddings, proposals, announcements of pregnancies, baby gender revelations, anniversaries and round-numbers birthdays. **Love** (*n*=191): this category included any text constituting a love declaration such as quotes from love songs or affectionate statements dedicated to partners, friends or family members. **No text** (*n*=149): this category included posts where the slogan of the cookies was not readable or cookies with logos rather than text. **Explicit content** (*n*=55): the slogan of the cookies of this category included explicit or sexual references. **Other** (*n*=884): is a final group which included posts in which the slogan on the cookies was not ascribable to any of the previous categories. Among others, it included external collaborations and partnerships, backstage moments of the company, special offers and discounts, and calls to action.

2.2. Online study

In order to analyze the posts at a finer level of granularity, we measured the funniness perception of the posts of the humorous subset and included the scores in the model as a covariate. For this purpose, an online study was conducted where people were asked to rate the funniness of a list of 30 or 31 posts randomly selected from the humorous subset. Before proceeding with the study, a manual check of the posts was performed, which led to the exclusion of 30 duplicates that were not detected by the Python Image Library. At this stage, the final subset of humorous slogans consisted of 241 Instagram posts. With a total of eight lists of posts and 80 participants, the online study guaranteed the collection of 10 funniness ratings per post. Since most of the slogans of the cookies were in French, participants were required to have at least a French proficiency at B1 level. This is recognized as the minimum level where non-native people are able to recognize different expressions of emotions (CEFR self-assessment grid) and have the necessary linguistic, pragmatic and sociolinguistic knowledge that is required to understand humor (Shively, 2013). Funniness perception was measured based on a three items 5-point Likert scale ranging from 'strongly disagree' to 'strongly agree' that was constructed based on existing literature about humor. The items developed to measure humor appreciation were created on the basis of an existing item adapted from Binsted, Pain and Ritchie (1997) (i.e. 'I found it funny') and two items created on the basis of the study from Chapman and Chapman (1974) (i.e. 'It made me laugh', 'It made me smile') where they stated that laughter and smiling are signs of humor appreciation.

2.2.1. Humor

The humorous subset was chosen for three main reasons. First, it was the theme category with the highest number of posts. Second, it was the most homogeneous category. In fact, in comparison with categories such as 'emotional events of life' or 'love', the posts present a lower degree of variation, as they mainly differ for social presence level and actual text. The other two categories, instead, showed a high degree of variation by often displaying different elements in the background such as people of different age and race, and having texts related to very different events of life such as weddings, birthdays or proposals that could also confound people's engagement with the posts. Third, past research showed a positive relationship between the use of humor on brand-generated social media posts and their number of likes and comments (Malhotra, Malhotra & See, 2013; Lee, Hosanagar & Nair, 2018).

2.2.2. Procedure

Participants were recruited through snowball sampling via Facebook groups of university students majoring in French language or personal connections via social media such as Facebook, Instagram and LinkedIn, and Survey Circle. Participants accessed the study via a link or a QR-code. After giving their informed consent, they were asked some demographic questions. Additionally, non-native speakers of French were asked to self-evaluate their level of French by choosing which of the six statements each corresponding to a level of the CEFR described their level best. Next, participants rated their funniness perception of the list of posts they were assigned to. Finally, they had the option to leave a comment before reading the debriefing.

2.3. Time

Having to deal with real-world data, means having to deal with noise. The main source of noise in our experiment is that posts that are public for longer have a higher potential to collect likes and comments through time than more recent posts. Therefore, we calculated how many days were elapsed since the post had been published, and included this count as a covariate in our two models.

3. Statistical analysis and results

In order to test the proposed hypothesis first at a coarse-grained and then at a fine-grained level, the two models

presented were tested performing a two-way MANCOVA using IBM SPSS 24.

3.1. Model 1: general effect of social presence

Recall that this model includes social presence as independent variable, number of likes and comments as dependent variables, and the theme of the slogans of the cookies and time as covariates. The results of the MANCOVA show that the covariate time is significantly related to customer's engagement overall $F(2, 1741) = 810.38$, $p < .001$, Pillai's Trace = .482, partial $\eta^2 = .48$. Specifically, the covariate is significantly related both to the number of likes, $F(1,1742) = 51.82$, $p < .001$, partial $\eta^2 = .03$, and to the number of comments, $F(1,1742) = 1142.12$, $p < .001$, partial $\eta^2 = .40$. There was a statistically significant difference between the levels of social presence on the combined dependent variables after controlling for time, $F(4, 3484) = 3.29$, $p = .011$, Pillai's Trace = .008, partial $\eta^2 = .004$, showing that overall social presence has an effect on customers' engagement. However, there was no significant individual effect of social presence on number of likes after controlling for time, $F(2,1742) = 2.54$, $p = .079$ and no significant individual effect of social presence was found on the number of comments after controlling for time, $F(2,1742) = 0.14$, $p = .870$.

Interestingly, the results showed that there was a statistically significant interaction effect between social presence and theme of the text on the combined dependent variables after controlling for time, $F(20, 3484) = 1.91$, $p = .008$, Pillai's Trace = .022, partial $\eta^2 = .011$. Specifically, a significant interaction effect of social presence and theme of the text on the number of likes of the posts after controlling for time was found, $F(10, 1742) = 2.77$, $p = .002$, partial $\eta^2 = .016$. However, no significant interaction effect between social presence and theme of the text when controlling for time was found for the number of comments, $F(10, 1742) = 0.89$, $p = .546$, partial $\eta^2 = .005$. The simple effect analysis showed that there is a significant effect of social presence levels on the number of likes for the 'love' theme category, such that consistent to hypothesis 1a, high levels of social presence ($M = 1373.70$, $SD = 933.02$) lead to higher numbers of likes than medium ($M = 362.55$, $SD = 501.42$), $p = .016$ and low levels of social presence ($M = 237.40$, $SD = 481.26$), $p = .017$. Moreover, an effect consistent to hypothesis 1b and partially consistent to hypothesis1a was found for the 'other' theme category, such that high levels of social presence ($M = 1202.58$, $SD = 1047.27$) lead to higher numbers of likes than low levels of social presence ($M = 417.33$, $SD = 606.11$), $p = .004$, and medium levels of social presence ($M = 1035.16$, $SD = 1418.89$) lead to higher numbers of likes than low levels of social presence, $p < .001$. At a coarse grained level, the results provided partial support for the expected hypothesis, showing that the expectation that higher levels of social presence lead to higher customers' engagement compared to lower levels of social presence is true in terms of number of likes for some themes of the slogans. Consequently, present results suggest that social presence does not have an effect on customers' engagement alone, but is also dependent on the type of posts it is related to, calling for a more fine-grained approach which focuses on a specific category of slogans.

3.2. Model 2: social presence in humorous posts

Recall that this model includes social presence as independent variable, number of likes and comments as dependent variables and funniness perception of the posts and time as covariates. The results of the MANCOVA show that the covariate time is significantly related to customer's engagement overall $F(2, 235) = 82.78$, $p < .001$, Pillai's Trace = .413, partial $\eta^2 = .41$. Specifically, the covariate is significantly related to both the number of likes, $F(1, 236) = 163.13$, $p < .001$, partial $\eta^2 = .41$, and comments, $F(1,236) = 25.12$, $p < .001$, partial $\eta^2 = .10$. The covariate funniness perception is not significantly related to customers' engagement overall, $F(2, 235) = 2.65$, $p = .073$, Pillai's Trace = .022, partial $\eta^2 = .022$. However, it is significantly related to the number of likes, $F(1, 236) = 4.98$, $p = .027$, partial $\eta^2 = .02$, but not to the number of comments, $F(1, 236) = 0.42$, $p = .520$, partial $\eta^2 = .002$. Moreover, there was no statistically significant difference between the levels of social presence on the combined dependent variables after controlling for time and funniness perception, $F(4, 472) = 0.58$, $p = .675$, Pillai's Trace = .010, partial $\eta^2 = .005$. Social presence does not have an effect on customers' engagement with humorous posts when controlling for both time and funniness perception, thus at a fine-grained level, the results show no support for the expected hypothesis.

4. Discussion

Being based on real-world data, the analysis posed some methodological challenges related to the control of noise in data. A lot of caution has been paid to create a corpus that was as much controlled as possible. The choice of the brand was a first step in that direction. In fact, Shanty Biscuits is a relatively small business that produces cookies which visually vary only in relation to the text impressed on them, resulting in a fairly homogeneous Instagram feed. However, the setting of the posts showed also some degree of variation in the levels of social presence and in the slogans impressed on the cookies. Therefore, the posts were categorized on the basis of these two factors in order to isolate those variables. Moreover, in order to control for possible confounds at different levels of granularity, two models and two approaches were proposed.

4.1. Model 1

The analysis of model 1 showed that the theme of the slogans significantly impacts the effect of social presence on customers' engagement intended as number of likes, but not as number of comments, such that for two particular themes social presence has indeed an effect on social presence consistent to our proposed hypothesis, whereas for the other themes no significant effect was found. The

results of the simple effect analysis show that the effect of social presence on the number of likes is significant for the posts of the category 'love' and 'others'. More specifically, for the category *Love* it was found that high levels of social presence significantly lead to higher numbers of likes than medium and low levels of social presence. For the category *Other*, the results show that high levels of social presence lead to a higher number of likes than low levels and that medium levels lead to higher numbers of likes than low levels of social presence, showing in both cases a partial support for the expected hypothesis. Why we only find an effect for a subset of the categories, is a topic for future research, but as we have seen in the results for Model 2, the content of the slogan (in the case of Model 2: funniness) is significantly related to the number of likes, so we expect the same to be the case here.

No significant effect of social presence was found on the number of comments under the posts, suggesting that particular attention to the possible difference between likes and comments should be taken into account in future research. Past research already pointed out that liking a post is quicker and less committing than commenting a post (Antheunis, van Kaam, Liebrecht & van Noort, 2016), which makes comments less direct. Moreover, commenters can also respond to each other, making the comment count a noisier signal to measure engagement. Future research could further investigate this difference to understand which other factors need to be taken into account when analyzing customers' likelihood to comment a post.

4.3. Model 2

The analysis of model 2 showed that no significant effect of social presence was found at this further level of granularity. Considering that the funniness perception was not a significant covariate in the model, the results of the analysis of the humorous subset are consistent with the ones of model 1, where no significant effect of social presence on engagement was found for the humorous slogans. The decision of focusing on the humorous subset for the fine-grained approach was taken before conducting the analysis of model 1, based on reasons that have been already explained. The results of the simple effect analysis in model 1 suggest that probably it would have been interesting to apply the fine-grained approach to the subset of love-related posts, in order to establish the strength of the effect of social presence at this finer level of granularity. Therefore, we suggest future research to take this aspect into account.

5. Conclusion

Present research proposed two possible approaches in order to test the effect of social presence on customers' engagement with brand-generated social media posts and to eventually establish its strength, while guaranteeing high ecological validity. Through a case study of an existing brand, the proposed hypothesis was tested at two levels of granularity. The aim of the research was to contribute to the question of how to control for variation when it comes to analyzing corpora of social media posts. Showing that indeed at different levels of granularity the strength of the effect of social presence changes, present research suggests that social presence needs to be analyzed in combination with other factors in order to have more accurate estimates of its effect.

6. References

Antheunis, M., van Kaam, J., Liebrecht, C., & van Noort, G. (2016). Contentmarketing op sociale netwerksites: Een onderzoek naar gedrag en motivaties van consumenten.*Tijdschrift voor Communicatiewetenschap*, *44*(4), pp. 337-365.

Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces engage us: Photos with faces attract more likes and comments on instagram. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* pp. 965-974.

Binsted, K., Pain, H., & Ritchie, G. D. (1997). Children's evaluation of computer-generated punning riddles. *Pragmatics & Cognition*, *5*(2), pp. 305-354.

Chapman, A. J., & Chapman, W. A. (1974). Responsiveness to humor: Its dependency upon a companion's humorous smiling and laughter. The *Journal of Psychology*, *88*(2), pp. 245-252.

Council of Europe. (n.d.). Self-assessment grid - Table 2 (CEFR 3.3): Common Reference levels. Retrieved from https://www.coe.int/en/web/common-european-framework-referencelanguages/table-2-cefr-3.3-common-reference-levels-self-assessment-grid

Cyr, D., Head, M., Larios, H., & Pan, B. (2009). Exploring human images in website design: a multi-method approach. *MIS quarterly*, *3*(33), pp. 539-566.

Droulers, O., & Adil, S. (2015). Could face presence in print ads influence memorization? *Journal of Applied Business Research (JABR)*, *31*(4), pp. 1403-1408.

Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising content and consumer engagement on social media: evidence from Facebook. *Management Science*, *64*(11), pp. 5105-5131.

Malhotra, A., Malhotra, C. K., & See, A. (2013). How to create brand engagement on Facebook. *MIT Sloan Management Review*, *54*(2), pp. 18-20.

Pansari, A., & Kumar, V. (2017). Customer engagement: The construct, antecedents, and consequences. *Journal of the Academy of Marketing Science*, *45*(3), pp. 294—311.

Shively, R. L. (2013). Learning to be funny in Spanish during study abroad: L2 humor development. *The Modern Language Journal*, *97*(4), pp. 930-946.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. John Wiley & Sons.

Yoo, Y., & Alavi, M. (2001). Media and group cohesion: Relative influences on social presence, task participation, and group consensus. *MIS quarterly*, *25*(3), pp. 371—390.

The Speech Act of Apologising in Japanese Computer-Mediated Discourse: A Corpus-Assisted Approach

Eugenia Diegoli
University of Bologna
E-mail: eugenia.diegoli2@unibo.it

Abstract

The study investigates the use of three Japanese expressions commonly signalled as apologetic, namely *gomen* 'sorry (NON-POLITE)', *sumimasen* '(I'm) sorry (POLITE)' and *mōshiwake arimasen* '(I'm) sorry/I apologise (SUPER POLITE)', in the written, online Q&A forum Yahoo! Chiebukuro. Such expressions are analysed within a large amount of co-text in a corpus of spontaneous conversations collected by the author. Combining statistical techniques (concordance tool) and qualitative analysis (close reading), the pragmatic functions served in the data by the three lexemes were categorised. When the lexemes were intended as an apology, the apology strategies employed in their immediate co-text and playing a role in the realisation of the apology were investigated. Moreover, three devices frequently co-occurring with the lexemes were identified: the adverbs *chotto* 'a little' and *nanka* 'somehow' and the verb-ending form *-te shimau*. I argue that they modify the force of the speech act by carrying additional, relevant meaning and, eventually, favouring an indirect interpretation of the utterance.

Keywords: digital speech acts, apologies, Japanese, (im)politeness, corpus-assisted discourse studies, pragmatics

Overview

The aim of this study is twofold. First, to investigate the use of three Japanese expressions commonly signalled as apologetic, namely *gomen* '(I'm) sorry (NON-POLITE)', *sumimasen* '(I'm) sorry (POLITE)' and *mōshiwake arimasen* '(I'm) sorry/I apologise' (SUPER POLITE)', in the Japanese written Q&A website Yahoo! Chiebukuro 'https://chiebukuro.yahoo.co.jp/'. Such expressions are what Blum-Kulka and Olshtain define as illocutionary force indicating devices (henceforth IFIDs), i.e. explicit linguistic markers of pragmatic force which selects a routinised expression of regret (1984, 198). The IFIDs are investigated as parts of larger linguistic units and, when they are intended as apologies, the apology strategies employed in their immediate co-text are also categorised. Second, to analyse the discoursive effects of lexical and grammatical devices that frequently co-occur with the three IFIDs and function as (im)politeness markers, playing a role in the realisation of the speech act. The two research questions (RQs) are as follows: (RQ1) When apologising in Computer-Mediated Discourse (henceforth, CMD), what are the apology strategies used by Japanese speakers in the immediate co-text of IFIDs? (RQ2) What are the lexical and grammatical devices that in CMD frequently co-occur with the IFIDs and modify the force of the speech act?

1. Theoretical background

The present work partially relies on Brown and Levinson's (1987) view of politeness as a system of linguistic devices used to redress face threat. Although I am indeed aware of the shortcomings their model presents (e.g., their focus solely on the speaker [Eelen, 2001; Watts, 2003] and their notion of a strategic exploitation of linguistic strategies [Matsumoto, 1988; 1989; Ide, 1989]) and, in many respects, approving of this criticism, I believe that their model does provide a useful terminology for the analysis of linguistic interactions and (im)politeness-related phenomena. More specifically, I continue to rely on their distinction between negative and positive politeness and their assumption that, generally speaking, indirectness tends to enhance the degree of politeness in the expression of speech acts. With reference to the latter, it is my contention that generalisations about the notion of indirectness as a means to achieve politeness – or, indeed, as a means to mitigate impoliteness – are useful for analysing speech acts in Japanese CMD – but it may not be so for other languages or language varieties. This is why the present analysis continues to rely on these notions, whilst acknowledging that negative and positive politeness strategies often work in tandem and their distinction is not always unequivocal, and that there is no unidirectional relationship between politeness and indirectness.

A second key theoretical model which involves inferences and attempts to deal with the interpretation of utterances is Sperber and Wilson's Relevance Theory (1995), which focuses on the cognitive processes by which hearers recognise the intention – i.e., the relevant information – beyond the literal meaning. The notion of relevance will be very helpful in the analysis of the extra information conveyed by additional lexical and grammatical devices co-occurring with the IFIDs. Such devices may be omitted because the utterance already conveys sufficient information. They thus appear to break Grice's (1975) maxims of quantity (avoid unnecessary information) and manner (avoid ambiguity, be brief) and might lead the hearer to assume that the additional information has some kind of relevance to that specific situation.

Particularly in the context of Japanese (im)politeness markers, the body of corpus-driven discourse studies into this kind of linguistic resources is to date comparably small. So far, research is available mainly (but not exclusively) from Japan, where such devices are usually referred to as *hairyo hyōgen*, which literally means 'expressions of consideration'. Three (im)politeness markers have attracted focused attention in the literature, and they will be addressed also in the present study.

Yamaoka (2004) shows that the low-degree adverb *chotto* 'a little', which is the first device selected for the analysis, is often used in Japanese when criticising or rejecting a request to minimise the face-threat. The second device under analysis is the verb-ending form *-te shimau*, grammaticalised from the verb *shimau* 'put away'. This auxiliary verb is a polyfunctional word which was found to be used as a marker of aspect such as completion or to convey a modal meaning (Tamura, 2007). This latter use of *-te shimau* expresses the speaker's negative feelings and attitude toward the event described in the main clause. More specifically, it conveys the speaker's disappointment and regret towards an event seen as regrettable and which they assume to have no control over (Fauconnier, 2013). Finally, the discourse marker *nanka* 'like, somehow' was also observed in the corpus. It can be used to convey the referential meaning 'something (like), anything' or with no referential meaning to convey many different functions (Sato, 1997, cited in Murayama, 1999). Particularly significant for the present study are instances where *nanka* is used as a mitigator increasing the social distance between interactants, while conveying reluctance to express a given utterance (in our case, an apology) straightforwardly.

2. Corpus construction and analysis

Drawing from Corpus-Assisted Discourse Analysis (CADS, Partington *et al.*, 2013) the present study combines a discoursive approach with corpus linguistic tools for the collection and analysis of naturally occurring data. Unlike traditional corpus linguists, CADS researchers emphasise a balance between quantitative (statistical) and qualitative (traditional close reading) approaches. Continuing this trend, this work focuses on the production of apologetic expressions whilst acknowledging that their meaning in interaction strongly relies on contextual information: they are deictic forms, anchored to other variables and to their surrounding co-text. In this perspective, the present study did employ basic corpus linguistic tools, namely the free software BootCat (Baroni & Bernardini, 2004) and the corpus query system Sketch Engine (Kilgarriff *et al*., 2014) for the corpus collection and analysis respectively, but close-reading of the texts and manual annotation of pragmatic functions were carried out as well, bringing Corpus Linguistics and aspects of Discourse Analysis together.

First, a search-term specific or *topical* corpus (Taylor, 2015) that consists of only webpages containing Japanese apologetic lexemes, namely *gomen, sumimasen*, and *mōshiwake arimasen* was collected from the URL of the Internet Q&A forum Yahoo! Chiebukuro 'https://chiebukuro.yahoo.co.jp/'. Then, utterances containing the apologetic lexemes were identified and analysed within their wide co-text with the aim of exploring their role in context. The corpus is thus used as a resource for retrieving the most frequent realisations of the so-called apologetic lexemes in discourse units, allowing the researcher to access a large number of instances of usage without sacrificing important contextual features vital for discourse analysis. The collected corpus contains 253,577 tokens. Using the concordance tool, a total of 1424 occurrences of explicit apologetic lexemes were identified and downloaded in plain text format. Duplicates in the corpus were manually eliminated and an analysis was conducted on 385 occurrences of *gomen*, 324 expressions with *sumimasen*, and 338 instances of *mōshiwake arimasen*, for a total of 1,047 discourse units analysed combining statistical overview techniques (mainly the concordance tool) and qualitative analysis (traditional close reading), in order to identify each utterance as performing a specific pragmatic function.

3. Results

The results suggest that the expressions of apology chosen for this study are polyfunctional words used in my data to perform the following twelve actions, arranged in descending order of use-frequency: apology for an offence, first-order considerations, idiomatic expressions, request, attention-getting device, request, conversation opener, thanks, mitigation of impolite expressions, rejection of a request or an offer, demand for an apology, and, finally, rejection of an apology. Thus, the real communicative intent of Japanese IFIDs commonly signalled as apologetic can be something different from an apology, or indeed can be apologising plus other intentions. When the lexemes *gomen*, *sumimasen* and *mōshiwake arimasen* where intended as an apology (in 55.6%, 42.9% and 72.5% of occurrences respectively, for a total of 597 instances of apology), the study also investigates the different patterns of speech act realisation, more specifically the apology strategies employed in their immediate co-text (RQ1).

The analysis of how apologies are structured beyond the single clause revealed that the communicative intent or illocutionary force of utterances with IFIDs is strongly affected by the use (or the omission) of three morphosyntactic devices, namely the adverbs *chotto* 'a little' and *nanka* 'somehow' and the verb-ending form *-te shimau*. I will argue that these devices can function as (im)politeness markers. These arguments are developed by analysing multiple linguistic, multimodal and contextual variables that frequently co-occur with apologetic devices in CMD and play a role in the realisation of their pragmatic functions.

3.1. Apology strategies

With reference to RQ1, the analysis of linguistic and contextual variables in the immediate co-text of apologetic lexemes revealed that an expression of responsibility either preceding or following the lexeme is the most common pattern in data set, used in 375 (62.8%) apologetic utterances, followed by 'detached' apologies (i.e. apology where the lexeme alone constituted the utterance) and self-humbling, used in 106 (17.8%) and 64 (10.7%) of cases respectively. However, these values change considerably across the three lexemes, as illustrated in Figure 1. The total number of instances for each IFID is also indicated enclosed between parenthesis.

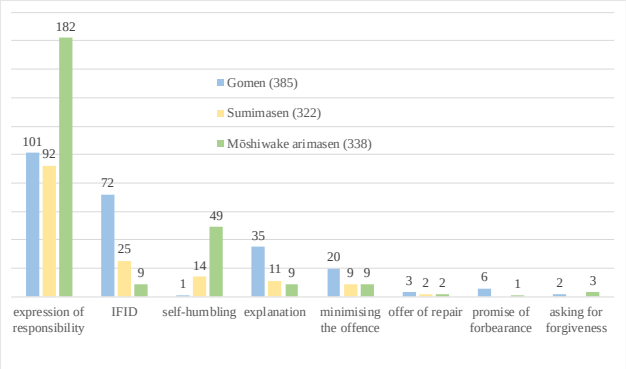


Figure 1: Apology strategies across the three IFIDs

Strategies involving the realisation of deference (i.e., admission of responsibility and self-humbling) appear to be comparatively quite frequent in formal settings, where *mōshiwake arimasen* (SUPER POLITE) is perceived as the appropriate form of apology. On the other hand, *gomen* (NON-POLITE) is much more common than its formal counterparts in formulaic apologies where there is no additional reference made in the clause to the offence (i.e., detached apologies) or no indication of regret on the speaker's part (i.e., minimising the offence). These findings support Brown and Levinson's assumption that negative politeness strategies involving deference are more common in asymmetrical situations where there is greater social and psychological distance between interactants (1987, 178).

A second finding is that, though strategies involving taking on responsibility are arguably more likely to be successful, there may be situations where the speaker is not willing to humiliate themselves and strategies minimising the offence are preferred. We can thus safely assess that the amount of face a speaker is willing to lose seems to play a role in the selection of apology strategies.

3.2. (Im)politeness markers

With reference to RQ2, the study also shows a total of 205 utterances containing what I identified as the three most frequent (im)politeness markers in the corpus, i.e. the lower-degree adverb *chotto* 'a little'; the verb-ending form *-te shimau*; and, ultimately, the adverb *nanka* 'like, somehow'. The identification of (im)politeness markers in the data is frequency-based because, as Terkourafi and Kádár (2017) have pointed out, an expression can be considered to be conventionalised to achieve a particular illocutionary goal only if it used frequently enough in that context. Figure 2 gives the relative distribution of IFIDs in co-occurrence with the three (im)politeness markers.

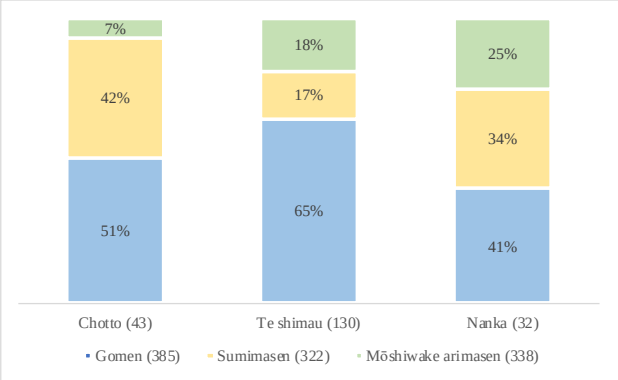


Figure 2: Percentages of co-occurrence of (im) politeness markers and IFIDs

It should be noted that the above figure represents patterns of co-occurrence as observed in a concordance with 100 characters of co-text on each side. It thus comprehends also instances where the three expressions investigated convey their primary meanings, which are not related to (im)politeness. Generally speaking, however, it is possible to state that the more formal the IFID, the less frequent (im)politeness markers are. This pattern is particularly clear for the adverb *chotto* and the verb-ending form *-te shimau*. These two devices co-occur quite frequently with *gomen* (NON-POLITE), while are less common with *sumimasen* (POLITE) and *mōshiwake arimasen* (SUPER POLITE).

When used in the L1 position (i.e., immediately at the left of the IFID), *chotto* signals that the utterance is not to be taken as an apology, but rather as a request or as an attention-getting device. According to Brown and Levinson (1987), a request is a speech act potentially threatening the addressee's negative face and, as such, it requires the use of mitigating devices conveying that no harm is intended. The set *chotto*+IFID is in fact generally used by Yahoo! Chiebukuro users to hedge the illocutionary force of the utterance with the aim to avoid a direct request, thus downgrading the force of the speech act.

The verb-ending form *-te shimau* is also employed as a device increasing the politeness of the utterance. It expresses the speaker's negative feelings and attitude toward the event described in the main clause – usually an event somehow related to the speaker and considered to be unwanted by the addressee, which thus requires an apology. By conveying that the addressee's wants are similar to the speaker's wants, it shows empathy to the addressee and boosts the interactants' positive faces. The use of *-te shimau* in apologies can therefore be considered as a positive politeness strategy serving a similar function to English intensifying devices such as *really*, *very*, etc.

Finally, the adverb *nanka*, when used in the L1 position with no referential meaning, depending on the interactional contexts can function as a marker of mock politeness, i.e. "the use of politeness strategies which are obviously insincere" (Culpeper *et al*. 2003, 1555). Generally speaking, it conveys reluctance to express a given utterance (in our case, an apology) straightforwardly. For

instance, it was found to introduce demanded apologies clearly felt as unnecessary by the speaker. Unlike *-te shimau*, it decreases the illocutionary force of the utterance, frequently overlapping with strategies minimising the offence. I argue that *nanka* in combination with apologetic IFIDs favours an indirect interpretation of the apologetic utterance, which carries additional meaning the hearer must assume to be relevant in that specific situation. The relevance of the additional information conveyed by *chotto* and *nanka* in co-occurrence with IFIDs seems to break Grice's (1975) maxims of quantity (avoid unnecessary information [45]) and manner (avoid ambiguity, be brief [46]), thus arousing Yahoo! Chiebukuro users' suspicion who discuss their use in the corpus. Such discussions provide emic, metalanguage comments supporting the interpretation given above.

The findings also reveal CMD-specific usage of multimodal features such as emoji, *kaomoji* (i.e., Japanese emoticons), non-standard script choices and elements orthographically mimicking phonetic processes. They generally boost interactants' positive faces by assuming common ground and decrease the perceived psychological and social distance with the addressee, as in the following example:

1. *Setsumei ga umaku kakenakute mōshiwake arimasen (kaomoji) shokuba de todokede o shinakereba naranai koto ga ari, senjitsu genpon 1 mai to sono kopī o sōfu shita tokoro henrei sarete shimaimashita.*

   説明が上手く書けなくて申し訳ありません。m (_ _) m 職場で届出をしなければならない事があり、先日原本1枚とそのコピー1枚を送付したところ返戻されてしまいました。

   I apologise for not being able to explain it very well (*kaomoji*) There was something I had to report at work, and a few days ago the original document and a copy that I had just sent returned back.

In the above excerpt, we can observe the use of the *kaomoji* m (_ _) m, which visually conveys the body movement of 'bowing'. Though it is usually associated with casual speech, it is employed here in an utterance where the dominant speech level is super polite. I argue that, in this specific context, it decreases the perceived distance between interactants. It also shows that in spontaneous conversations negative (formal speech) and positive (multimodal features, but also *-te shimau* at the end of the utterance) politeness work together and their distinction is more often than not rather blurred.

4. Conclusion

This paper has dealt with IFIDs, namely formulaic expressions of apology, in CMD and, more specifically, in the Japanese Q&A website Yahoo! Chiebukuro. The polyfunctionality of IFIDs has been addressed and focused attention has been given to morphosyntactic and discoursive devices co-occurring with IFIDs and eventually affecting the illocutionary force and/or the politeness level of the utterance. With regard to the realisation of apologies in CMD, eight main apologetic strategies have been classified. Their distributions across the three IFIDs supports Brown and Levinson's (1987) assumption that strategies involving deference (negative politeness strategies) are perceived as appropriate in formal settings, which possibly requires a greater face-loss on the speaker's part for the apology to be successful. On the other hand, in informal settings instances of strategies minimising the responsibility for an offence were also found, showing that there is a limit to the amount of face a speaker is willing to lose when apologising. We can thus asses that the use of IFIDs specialised for the expression of speech acts commonly considered as 'polite' does not necessarily produce polite utterances. What seems to be more relevant is whether the IFID is accompanied by what we have called "(im)politeness markers". When this is the case, such additional forms can modify the IFID and produce marked utterances that may be perceived as polite or non-polite/impolite depending on the interactional context and surrounding co-text.

The analysis also revealed multimodal, CMD-specific features of speech acts such as *kaomoji*. They generally boost interactants' positive faces by assuming common ground and decrease the perceived psychological and social distance with the addressee. An additional finding is that multimodal features such as *kaomoji*, though still more common in informal discourse, were commonly found in formal speech as well, possibly signaling that this kind of visual resources is becoming more and more pragmatically unmarked.

5. References

Baroni, M., and Bernardini, S. (2004). BootCaT: Bootstrapping corpora and terms.

Blum-Kulka, S., and Olshtain, E. (1984). Requests and Apologies: a cross-cultural study of speech act realisation patterns (CCSARP). *Applied Linguistics*, 5 (3), 196-213.

Brown, P., and Levinson, S. C. ([1978]1987). *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Elen, G. (2001). *A critique of politeness theories*. Manchester: St. Jerome Pub.

Fauconnier, S. (2013). Completives as markers of non-volitionality. In *Folia Linguistica*, 47, 35-53.

from the web. Proceedings of LREC (2004). Available at: http://www.lrec-conf.org/proceedings/lrec2004/pdf/509.pdf (retrieved: 2020/09/27).

Grice, H. P. (1975) Logic and conversation. In P. Cole and J. Morgan (Eds.) *Studies in Syntax and Semantics III: Speech Acts*, New York: Academic Press, pp. 183-98.

Ide, S. (1989). Formal forms and discernments: Two neglected aspects of linguistic politeness. *Mulitlingua*, 8, 223-248.

Kilgarriff, A., Baisa, V., Bušta, J, Akubíček M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography ASIALEX*, 1. DOI: 10.1007/s40607-014-0009-9.

Matsumoto, Y. (1988). Reexamination of the Universality of Face: Politeness Phenomena in Japanese. *Journal of*

Pragmatics, 12, 403-426.

Matsumoto, Y. (1989). Politeness and conversational universals: observations from Japanese. *Multilingua*, 8(2-3), 207-222.

Murayama, E. (1999). *The Multiple Functions of the Discourse Marker Nanka in Japanese Conversation*. Doctoral thesis. University of Hawaii, Manoa.

Partington, A., Duguid, A., and Tayor, C. (2013). *Patterns and Meaning in Discourse*. London: Routledge.

Sperber, D., and Wilson, D. ([1986]1996). *Relevance: communication and cognition*. 2a ed. Oxford: Blackwell.

Tamura, T. (2007). A note on the modal meaning of V-te shimau. In *Tsukuba English Studies*, 26, 11-33.

Taylor, C. (2015). *Mock Politeness in English and Italian: A Corpus-assisted Study of the Metalanguage of Sarcasm and Irony* (Unpublished doctoral dissertation). Lancaster University, Lancaster, United Kingdom.

Usami, M. (2002). *Discourse politeness in Japanese conversation: some implications for a universal theory of politeness*. Hitsuji Shōbō: Tokyo.

Watts, R. (2003). *Politeness*. Cambridge: Cambrdige University Press.

Yamaoka Masari (2017). Hairyo hyōgen no kanshūka o meguru - kōsatsu - metafā to no anarojī o moto ni (On the conventionalisation of considerate expressions – considerations on metaphor and analogy). In *Nihongo nihon bungaku*, 27-38.

Referring to Other Participants in Asynchronous Online Discussions: Citation Patterns in a Higher Education Context

Francisco Javier Fernández-Polo, Mario Cal-Varela
University of Santiago de Compostela
E-mail: xabier.fernandez@usc.es, mario.cal@usc.es

Abstract

The effectiveness of asynchronous online discussions as a learning tool in higher education depends critically on the participants' ability to create a sound social space (Swan, 2002). Referring to one another's messages is one way to display a sense of affiliation which strengthens bonds in the learning community. However, research has shown that students often fail to exploit this dimension of the forum (Murphy, 2004; Wise et al., 2012).

In this paper we look at references to previous posts in a set of online discussions held during a one-term undergraduate course. For the purposes of this analysis, sixteen full discussions —885 messages from 121 different participants in total— were drawn from the SUNCODAC corpus (Cal Varela & Fernández Polo, 2020). We describe the distribution of these references and analyse their degree of personalization and the kind of stance expressed, and explore how these issues might be affected by aspects of the students' identities such as gender and lingua-cultural background.

Our data reveal lower interactivity than expected. Almost half the posts analysed contain no references to previous contributions, exchange students being especially reluctant to mention their classmates' messages. The overall pattern is one of slight decline in the number of references over time, although interestingly during the second half of the course a small group of students increase the frequency of mentions of previous posts and they also tend to make these references more personal. The preferred practice is using first names, although the more formal name+surname pattern becomes more frequent during the second half of the course, which suggests increasing awareness of the academic character of the activity. Finally, simple acknowledgement is by far the most frequent form of alignment, and about a third of the contributions contain expressions of agreement. This seems to indicate that in most cases the main function of these references is preserving social relationships and avoiding conflict rather than engaging critically with differing views.

Keywords: computer-mediated communication, online discussion, personal references

References:

Cal Varela, M., & Fernández Polo, F. J. (2020). SUNCODAC: A Corpus of Online Forums in Higher Education. *Nexus*, *2*, 44–52. https://aedean.org/wp-content/uploads/nexus-2020-n02.pdf

Murphy, E. (2004). Recognising and promoting collaboration in an online asynchronous discussion. *British Journal of Educational Technology*, *35*(4), 421–431. https://doi.org/10.1111/J.0007-1013.2004.00401.X

Swan, K. (2002). Building learning communities in online courses: The importance of interaction. *Education, Communication & Information*, *2*(1), 23–49.

Wise, A. F., Perera, N., Hsiao, Y.-T., Speer, J., & Marbouti, F. (2012). Microanalytic case studies of individual participation patterns in an asynchronous online discussion in an undergraduate blended course. *Internet and Higher Education*, *15*, 108–117.

Snap-by-Snap!
How to Analyze Identity Work of Young Female Footballers on Snapchat

Melanie Fleischhacker, Eva-Maria Graf
University of Klagenfurt
E-mail: melanie.fleischhacker@aau.at, eva-maria.graf@aau.at

Abstract

Social networking sites (SNS) allow for complex (inter-)actions between users and the medium by means of their specific technical affordances and limitations; this facilitates users' multimodal discursive construction of multifaceted, flexible, in-the-moment identities. With its image-based and ephemeral, time-limited content, Snapchat is particularly popular among younger people. Via Snaps, users creatively produce fluid and semiotically complex online identities that enter the social world through interaction with others. This project aims to carve out the multimodally created, managed and negotiated (youth, gender, athletic, cultural, etc.) identities of young female (Austrian) football players and to analyze how these interact with the medium, Snapchat. For this purpose, "My stories" from 7 players were collected during 5 months in 2018. Our contribution proposes a multi-perspective, multi-level and multimodal linguistic analysis involving both bottom-up and top-down methodological approaches, yet also documents and critically addresses the challenges of linguistically analyzing social practices such as Snaps.

Keywords: Snapchat, Snaps, (youth) identity construction, linguistic analysis, multimodality

1. Introduction

Computer-mediated communication (CMC) in general and social networking sites (SNS) in particular form complex "in-between" spaces, in which situation factors (i.e., users with various purposes, characteristics, topics, and identities) interact with a specific medium and its affordances and limitations (Herring et al., 2013; Herring, 2007). Based on these, each platform or medium creates unique conditions that allow for different forms of (inter-)actions via the use of (various) socio-semiotic practices. Such actions and practices associated with particular technologies are again "nestled with other cultural practices and together they make available recognisable social identities" (Jones et al., eds., 2015, 3), i.e., digital practices interact in complex ways with physical and virtual spaces, times, interaction orders, and cultures. Concurrently, they build on various semiotic resources and modes such as text, audio, visual (image, video, color, font, etc.) or layout (Jewitt ed., 2017; Kress & van Leeuwen, 2006).

The goal of this research project is to investigate – from a multimodal linguistic perspective – how young female (athletic) selves and a particular social medium (Snapchat) interact and how this enables the *pro*sumption of the users' multifaceted identities in flux (Buckingham, 2008; Zajc 2015). The analysis focuses on the football players' ludic-creative identity work as well as their co-construction and management of youth, gender, athletic, cultural, etc. identities. We target the specific affordances and constraints of Snapchat in the multimodal (self-) construction of personal, relational, and social identities-in-action. Moreover, we propose a multi-level and multi-perspective methodology and document the complexity as well as the methodological and theoretical challenges for such a multimodal linguistic analysis. In this way, we aim to contribute to the emerging body of linguistic research on multimodal meaning-making practices and identity construction on SNS.

2. Identity Construction on Social Media

In post-structuralist linguistics, identity is conceptualized as socially and interactionally constructed in and through language and other semiotic resources or modes (Bucholz & Hall, 2005). Virtual identities or identities online are thereby considered to exist alongside offline identities; online self-presentations play with and foreground particular aspects of an "authentic" offline identity (Tagg, 2015; Benwell & Stokoe, 2006). Thus, "stories told and the histories recreated online are shaped by other identity markers (i.e. sexuality, ability, nationality, etc.) all of which translate from face-to-face realities into virtual ones" (Roseboro, 2014, 360). Instead of surpassing (the constraining) real-world categories such as gender, race, or ethnicity, identity work online is thus still oriented along existing stereotypes, norms, and values in particular communities of practice (Mackenzie, 2021; Veum & Undrum, 2018; Eckert & McConnell-Ginet, 1992).

Offline and online identities are complex, fluid, and situated, i.e., a series of selves adapted to a specific context, medium, or genre (Buchholz & Hall, 2005). In this way, an individual's portfolio of identities is discursively produced and interactively negotiated with the means of various semiotic resources or modes all of which encompass affordances and limitations for identity work. Identities performed online, then, result from CMC qualities and technological properties (of the particular medium or SNS) and how these enable different types of social organization or group interaction (Tagg, 2015; Seargeant & Tagg, 2014).

3. Snapchat

Snapchat, launched in 2011 (Velten & Rauf, 2016), has accumulated more than 300 million users worldwide by now, most of them young adults and teenagers (Statista, 2021). Users create and exchange 3 billion Snaps and produce around 400 million stories each day (Dunn & Langlais, 2020; Page, 2018). Apart from allowing Snapchat users to send (synchronous) private visual messages and short videos ("Snaps"), similar to other (instant) messaging services, the "My Story" feature lets users share Snaps with their (larger) social network (i.e.,

friends) for a maximum of 24 hours. This ephemerality of Snaps (disappearing, temporally limited and self-destructive content, with no persistent archive) represents the most salient feature of Snapchat (Piwek & Joinson, 2016). Additionally, videos are limited to 10 seconds and included text to 50 characters. Also, "no specific affordances for aggregated social feedback" (Bayer et al., 2015, 959) (such as likes, comments etc.) are provided. Snaps may be composed of various modes and semiotic resources (text, image, emoji, sticker, audio, video, drawing), but are primarily image-based. According to Page (2018), due to a "camera-first technology … the creation of visual content precedes the ways in which users can shape that content" (80). This embodies the underlying philosophy of the medium, i.e., sharing immediate, in-the-moment experiences and the specific visual meaning created in these images (Zappavigna, 2016).

In this way, Snaps function as digital practices and discourse that enable identities to enter the social world through interaction with others (Page, 2018; Jones et al., 2015; Bucholtz & Hall, 2005), i.e., via Snaps and the various resources and modes employed to create them, users construct (online) identities as well as personal relationships.[1]

At this point, Snapchat is still a relatively under-researched social messaging app or media-sharing platform. Existing research focuses primarily on user motivation and experience (e.g., Bayer et al., 2015; Piwek & Joinson, 2016) and on its influence on interpersonal relationships (Vaterlaus et al., 2016). Apart from Page's research (2018) on intersubjectivity, defined as the levels through which the producer of a Snap mediates their perspective on the viewed content, no linguistic research on Snapchat is available to date. Moreover, studies looking into the multimodal construction of various identities on social media are generally rare. In this way, the envisioned research project addresses existing research gaps – in the vein of Veum and Undrum (2018) for Instagram – by adding to the emerging body of linguistic literature on multimodal meaning-making and identity construction.

4. Purpose and Research Questions

MacIntosh et al. (2012, 213) describe the Internet as an interstitial space where "youth engage with, share, and construct identificatory, sociocultural, political, informational, and other frameworks". Young people spend more time with technology than any other daily activity (Coyne et al., 2013.), and they represent the majority of Snapchat users worldwide. Based on its creative possibilities for identity construction and relationship building, as well as young people's extensive use, Snapchat is a promising epistemological site for the analysis of the following research questions:

- What are recurring multimodal practices of identity construction on Snapchat?
- What kinds of (individual and collective) identities do the users construct or foreground via their Snaps?
- How do the technical affordances and limitations of Snapchat interact with the various identities that participants (un)intentionally stage?

5. Proposed Methodology

To carve out the various user identities and (multimodal) practices used to construct them and to critically relate these to underlying socio-semiotic/-cultural processes, we propose a multi-level, multimodal and multi-perspective linguistic analysis. Our analysis is located both in social semiotic multimodality and its focus on the choices users make (from the resources available to them on Snapchat) as well as in multimodal discourse analysis and its focus on the semiotic meta-functions the various modes in isolation as well as in combination fulfill (Veum & Undrum, 2018; Jewitt ed., 2017). Moreover, we adapt the text-based approach of Critical Discursive Psychology (Wiggins, 2017) and its analytical concepts subject positions, interpretative repertoires, and ideological dilemmas to our multimodal data in order to critically assess users' online identification processes (for a similar analytic endeavor see Fleischhacker, 2019). Analysis consists of three consecutive, though interconnected and mutually informed, steps and is carried out with the help of Atlas.ti, followed by a top-down socially informed interpretative step.

5.1. Data and Participants

Data was collected as part of a larger research project involving the U16 girls' football team of the First Vienna Football Club in Austria in 2018. Of the entire squad, seven players gave consent to the collection of social media data. All the participants were between 14 and 16 years of age, lived in Vienna and were experienced football players. Their social, ethnical, and national backgrounds were diverse, so were their native languages.

Figure 1: Example Series (combination of 4 Snaps)

Between April and August 2018 all "My Stories" published by these female players were collected. During the time of data-collection, audio-visual recording of videos was not

[1] For the purpose of this analysis, discourse is perceived as "the ways people build and manage their social worlds using various semiotic systems" (Jones et al., 2015, 4).

possible due to technical restrictions and the ephemerality of the data itself. Screenshots were taken of each story (i.e., one-to-many, asynchronous, semi-private Snaps), including multiple screenshots (i.e., stills) of moving Snaps to capture visual changes. Audio, however, was not considered. The entire corpus consists of 239 items, whereby we differentiated between individual Snaps (single images), series (a number of thematically or otherwise linked single images, which can incorporate different modes), and sequences (stills of moving images).

5.2. Levels of Analysis

To answer our research questions, three interrelated levels of analysis seem necessary.

(1) Focus on the textual function of Snaps: Modes (i.e., inventory of semiotic resources), their frequencies, and how (often) they are combined by users to create Snaps are analyzed using multimodal social semiotics (Kress & van Leeuwen, 2006). Modes considered for this research include image, written text, emojis, stickers, figures, filters or elements of altered reality as well as drawings. At this point, more than 30 possible combinations of modes that together form unique "multimodal ensembles" (Jewitt, 2009) have been identified in the data. These multimodal combinations are examined in terms of the meaning created between them (alignment, complementation, contradiction) (Jewitt, 2009). Since image-text combinations are frequent (despite the camera-first technology (Page, 2018)), the status between images and written text (i.e., as equal or unequal in terms of conveying the overall message of the Snap) is further detailed by adapting Martinec and Salwey's (2005) system for image-text relations.

(2) Focus on the ideational and interpersonal function: The ideational and interpersonal functions of Snaps expressed with the help of a complex combination of modes are then analyzed using Halliday's Systemic Functional Grammar (1985) and multimodal social semiotics (Kress & van Leeuwen, 2006), with a primary focus on images and written text. In line with Page's (2018) analysis of Snapchat, we examine the ideational function (i.e., what the Snaps represent) and focus on (visual) content or themes (such as football training or matches, self- and other-presentation, school and leisure activities, moods, and feelings etc.) and speech acts (see Engel, 1996). An analysis of the interpersonal function (i.e., how Snaps construe relationships between producer, viewer, and object) considers the presence of the Snap creator, the relationship between people represented in the Snap and their viewers (applying distance, address, and angle as suggested in van Leeuwen's (2008) "Representation and Viewer Network":), as well as the relationship between depicted actors (applying number of actors involved in action (or not) as suggested in van Leeuwen's (2008) "Visual Social Actor Network") and their proximity.

(3) Focus on users' identity construction and embedded ideologies: Finally, to investigate how the participants construct their various identities with the help of multimodal practices, an adaptation of Critical Discursive Psychology (CDP) (Edley, 2001; Wetherell, 1998) that includes visual and multimodal resources is necessary. CDP's analytical tools include *subject positions* (used to index varying and situated self- and other-positionings), *interpretative repertoires* (flexible resources or discursive patterns to talk about topics, issues, and events) and *ideological dilemmas* (competing arguments that emerge over time based on diverging interpretative repertoires). With the help of multimodal discourse analysis, then, individual practices of identity construction are related to broader discourses and ideologies in the context of youth, social media, and digital communication.

By combining these three levels of analysis (from the basic semiotic components of Snaps to multimodal ensembles and their functions, to the various identities constructed by these users and their socio-cultural embeddedness) and methodological approaches (multimodal social semiotics, SFL, CDP, multimodal discourse analysis), the descriptive categories can be interpreted in relation to the underlying "social fabric" (Wetherell, 1998, 403), while a top-down, interpretative approach is informed by a micro-level analysis of smaller constituents. In this way, the linguistic micro-level can be (critically) related to the (ideological) patterns of collective sense making (Wetherell and Edley, 1999), thus rendering more complex and varied, complementary results (Fleischhacker, 2019).

6. Methodological Challenges and Future Directions

The Snapchat data provide socio-culturally and technologically up-to-date insights into (online) identity constructions of young female athletes (in Austria). Yet, upon developing and (preliminarily) applying the proposed methodology, certain challenges have surfaced in carving out these multi-layered and multimodal online identities.

First, while researchers are still seeking for robust and replicable methods to make sense of multimodal data (Mondada, 2019; Page, 2018), Snapchat data is particularly complex and demanding to analyze. A series (see Figure 1), for instance, combines a number of Snaps with several modes in various combinations and with different relationships between them. Additionally, participants make use of multimodal practices such as re-contextualisation, embedding, etc. Ideally, for each mode, the ideational and interpersonal functions should be considered to eventually decode the combined message and function of the entire item. However, such functions have yet to be established for modes such as emojis, stickers or – even more challenging – drawings. According to Flewitt et al. (2017) the complex simultaneity of different modes, their structure, and materiality has not yet been resolved when attempting a (simultaneous) reading or interpretation.

Furthermore, without direct access to the process of our users' Snap creation or community of practice, various analytical challenges arise. Which modes were, for instance, actively used in a Snap that is based on a screenshot? At what point should multiple uses of modes

be coded in a Snap? How to reconstruct the combination of modes in Snaps that are based on screenshots or recontextualised images? How can the meaning of Snaps with a restricted audience design be reconstructed?

Second, the employed modes and their multimodal combinations are the building blocks for visual, textual, and otherwise constructed identities, to be analysed via interpretative repertoires and subject positions. Such an analysis of the participants' identities, however, involves an adaptation of the tools employed by CDP to include other semiotic resources and modes apart from language (and determining how these represent e.g., gendered interpretative repertoires) (see Fleischhacker, 2019).

Third, although methodological transparency, replicability, and academic rigor are to be pursued at all times, developing a coding manual to achieve intercoder reliability for such a complex analysis seems unfeasible. What is more, given the need for a certain familiarity with the football players and their situations to make sense of the Snaps, researchers unfamiliar with the participants would inevitably encounter problems, even when provided with extensive coding manuals.

Fourth, combining bottom-up / descriptive approaches and top-down / interpretative linguistic analysis also involves a synthesis of different (and potentially opposing) theoretical underpinnings and approaches to discourse. Structural and functionalist (i.e., post-structuralist) paradigms focus on the units or constituents of discourse (eliding the context), on the one hand or the socio-cultural practices and meanings that inform discourse (focusing less on structures and regularities), on the other; but they do not usually combine both (Schiffrin, 1994). Though Wetherell (1998), for instance, perceives a synthesis of micro- and macro-level analyses as fruitful and productive, others assume a critical stance to such a proposition, particularly in terms of the analytical context and the role of the researcher (Schegloff, 1997). However, a combined approach would respond to and embrace the idea that *discourse* involves both linguistic and social resources and unite them in analysis and interpretation (Schiffrin, 1994). Addressing these methodological (and theoretical) issues will provide (linguistic) researchers beyond the current project with necessary tools and valuable insights to address research questions in the context of social media corpora, based on a combination of both bottom-up and top-down approaches.

7. References

Bayer, J., Ellison, N., Schoenebeck, S. and Falk, E. (2015). Sharing the small moments: Ephemeral social interaction on Snapchat. *Information, Communication & Society,* 19(7), pp. 956-977.

Benwell, B., Stokoe, E. (2006). *Discourse and identity.* Edinburgh: Edinburgh University Press.

Bucholtz, M., Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies,* 7(4-5), pp. 585-614.

Buckingham, D. (2008). Introducing identity. In D. Buckingham (Ed.), *Youth identity, and digital media*. Cambridge, MA: The MIT Press, pp. 1-24.

Coyne, S. M., Padilla-Walker, L. M. and Howard, E. (2013). Emerging in a digital world: A decade review of media use, effects, and gratifications in emerging adulthood. *Emerging Adulthood*, 1, pp. 125-137.

Dunn, T.R., Langlais, M.R. (2020). 'Oh Snap!': A mixed-methods approach to analyzing the dark side of Snapchat. *The Journal of Social Media in Society*, 9(2), pp. 69-104.

Eckert, P., McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology*, 21, pp. 461-490.

Edley, N. (2001). Analysing masculinity: Interpretative repertoires, ideological dilemmas and subject positions. In M. Wetherell, S. Taylor & S.J. Yates (Eds.), *Discourse as data: A guide for analysis*. London: SAGE, pp. 189-228.

Engel, U. (1996). *Deutsche Grammatik*, 3rd ed. Heidelberg: Julius Groos Verlag.

Fleischhacker, M. (2019). Multimodal discourses of gender and sexuality in Austrian EFL textbooks. MA Thesis. University of Klagenfurt.

Flewitt, R., Hampel, R., Hauck, M. and Lancaster, L. (2017). What are multimodal data and transcription? In C. Jewitt (Ed.), *The Routledge Handbook of Multimodal Analysis*. Abingdon: Routledge, pp. 44-60.

Halliday, M. (1985). *An Introduction to Functional Grammar.* London: Arnold.

Herring, S. (2007). A faceted classification scheme for computer-mediated discourse. *Language@Internet,* 4 article 1. Accessible online via: https://www.languageatinternet.org/articles/2007/761

Herring, S., Stein, D. and Virtanen, T. (Eds.). (2013). *Pragmatics of computer-mediated communication*. Berlin: de Gruyter. (Handbooks of Pragmatics, Vol.9).

Jewitt, C. (2009). An Introduction to Multimodality. In C. Jewitt (Ed.), *The Routledge Handbook of Multimodal Analysis*. Abingdon: Routledge, pp. 14-27.

Jewitt, C. (Ed.). (2017). *The Routledge Handbook of Multimodal Analysis*, 2nd ed. Abingdon: Routledge.

Jones, R., Chik, A. and Hafner, C.A. (2015). *Discourse and digital practices. Doing discourse analysis in the digital age*. Abingdon: Routledge.

Kress, G., van Leeuwen, T. (2006). *Reading Images. The Grammar of Visual Design*, 2nd ed. Abingdon: Routledge.

MacIntosh, B.L., Poyntz, S. and Bryson, M.K. (2012). Internet. In N. Lesko & S. Talburt (Eds.), *Keywords in youth studies. Tracing affects, movements, knowledge*. New York: Routledge, pp. 213-218.

Mackenzie, Jai. (2021). Analysing gendered discourses online: child-centric motherhood and individuality in Mumsnet Talk. In J. Angouri & J. Baxter (Eds.), *The Routledge Handbook of Language, Gender, and Sexuality*. Abingdon: Routledge, pp. 408-422.

Martinec, R., Salway, A. (2005). System for image-text relations in new (and old) media. *Visual communication*,

4(3), pp.337-371.
Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics*, 145, pp. 47-62.
Page, R. (2018). Group selfies and snapchat: From sociality to synthetic collectivisation. *Discourse, Context & Media,* 28, pp. 70-92.
Piwek, L., Joinson, A. (2016). 'What do they Snapchat about?' Patterns of use in time-limited instant messaging service. *Computers in Human Behaviour,* 54(2016), pp. 348-367.
Roseboro, D.L. (2014). Mediated youth, curriculum, and cyberspace: Pivoting the in-between. In A. Ibrahim & S.R. Steinberg (Eds.), *Critical Youth Studies Reader*. New York: Peter Lang, pp. 360-369.
Schegloff, E.A. (1997). Whose Text? Whose Context? *Discourse & Society* 8, pp. 165-187.
Schiffrin, D. (1994). *Approaches to discourse*. Oxford: Blackwell Publishing.
Seargeant, P., Tagg, C. (Eds.) (2014). *The Language of Social Media*. Basingstoke: Palgrave Macmillan.
Statista. (2021). Anzahl der täglich aktiven Nutzer von Snapchat nach Regionen weltweit vom 1. Quartal 2014 bis zum 1. Quartal 2021. Accessible online via: https://de.statista.com/statistik/daten/studie/553714/umfrage/anzahl-der-taeglich-aktiven-snapchat-nutzer-nach-regionen/
Tagg, C. (2015). *Exploring digital communication. Language in action.* Abingdon: Routledge.
Van Leeuwen, T. (2008). *Discourse and practice. New tools for Critical Discourse Analysis*. Oxford: OUP.
Vaterlaus, J.M., Barnett, K., Roche, C. and Young, J.A. (2016). 'Snapchat is more personal': An exploratory study on Snapchat behaviors and young adult interpersonal relationships. *Computers in Human Behavior,* 62(2016), pp. 594-601.
Velten, J., Rauf, A. (2016). The influence of Snapchat on interpersonal relationship development and human communication. *The Journal of Social Media in Society,* 5(2), pp. 5-43.
Veum, A., Undrum, L. (2018). The selfie as global discourse. *Discourse & Society,* 29(1), pp. 86-103.
Wetherell, M. (1998). Positioning and interpretative repertoires: Conversation Analysis and Post-structuralism in dialogue. *Discourse and Society,* 9(3), pp. 387-412.
Wetherell, M., Edley, N. (1999). Negotiating hegemonic masculinity: Imaginary positions and psycho-discursive practices. *Feminism & Psychology,* 9(3), pp. 335-356.
Wiggins, S. (2017). *Discursive Psychology. Theory, method and applications*. London: SAGE.
Zajc, Melita. (2015). Social media, presumption, and dispositives: New mechanisms of the construction of subjectivity. *Journal of Consumer Culture,* 15(1), pp. 28-47.
Zappavigna, M. (2016). Social media photography: Construing subjectivity in Instagram images. *Visual Communication,* 15(3), pp. 271-292.

Syntactic Variation and Interactional Coherence in Online Communication: The German Conjunction *weil* in Written Interactions

Aivars Glaznieks
Institute for Applied Linguistics, Eurac Research Bolzano
E-mail: aivars.glaznieks@eurac.edu

Abstract

The important role of conjunctions in computer-mediated communication (CMC) was first highlighted by Baron (2010). She focused on so-called *utterance break pairs* and analysed the linguistic means used to signal cohesion between the utterance chunks in Instant Messages. According to Baron's findings, conjunctions are relevant for creating coherence not only within a single text, but also between utterance pairs. According to an observation made by Günthner (2012), utterance break pairs in CMC are sometimes jointly produced by chat partners, e.g. when one chat partner expands on a previous message of another person. In conversation analysis, the term *collaborative production* is used for an utterance that is not produced by a single person but jointly by at least two.

In the present study, I will introduce some findings of a project that investigated the use of conjunctions in the German Facebook corpus DiDi (Frey et al., 2016). I will focus on the use of the conjunction *weil* ('because') in German Facebook texts and analyse how it is used as a cohesive device in cross-message usages such as question-response pairs, utterance break pairs and collaborative productions with respect to word order variation in contemporary German. In Standard German, *weil* is used as a subordinating conjunction, which causes verb-final word order. In spoken language, *weil* is also frequently used as a coordinating conjunction, causing verb-second word order. Regarding CMC, recent studies show a clear preference for verb-final word order, e.g. on Wikipedia discussion pages (Storrer, 2019).

The results of our study show that subordinate *weil*-clauses are used in both utterance break pairs and in collaborative productions whereas coordinate *weil*-clauses are used in utterance break pairs of a single user. Thus, word order variation can be used to disambiguate reference in chats, particularly regarding non-adjacent utterance break pairs.

Keywords: interactional coherence, conjunctions, social network sites, German CMC

References

Baron, N. (2010). Discourse Structures in Instant Messaging: The Case of Utterance Breaks. *Language@internet,* 7, article 4.

Frey, J.-C., Glaznieks, A. and Stemle, E.W. (2016). The DiDi Corpus of South Tyrolean CMC Data: A multilingual corpus of Facebook texts. In A. Corazza, S. Montemagni & G. Semeraro (Eds), *Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016. 5-6 December 2016, Napoli*. Torino: Academia University Press, pp. 157-161.

Günthner, S. (2012). „Geteilte Syntax": Kollaborativ erzeugte dass-Konstruktionen. *gidi Arbeitspapierreihe,* 43 (08/2012).

Storrer, A. (2019). Text und Interaktion im Internet. In L. Eichinger & A. Plewnia (Eds), *Neues vom heutigen Deutsch*. Berlin: de Gruyter. pp. 221–244.

Sustainability Communication of Nestlé on Twitter in the German and French Context

Clarissa Glück, Marie Steffens
Utrecht University
E-mail: clarissa.glueck@t-online.de, m.g.steffens@uu.nl

Abstract

This paper examines the differences in sustainability communication of the two Twitter profiles, Nestlé Germany and Nestlé France. The world's largest food company Nestlé is faced with the challenge of maintaining legitimacy towards international stakeholders through sustainability issues. The question arises whether the company is using the theory of cultural dimensions, which is widespread in management (Hofstede et al., 2010), in order to communicate sustainability in international markets in a target group-oriented manner. This also would be the case for the important European markets Germany and France. Additionally, the different classification of the two countries in the Food Sustainability Index 2019 raises the question of whether food sustainability is weighted differently in the two countries and whether communication about it is different.

Keywords: sustainability communication, Twitter, cultural dimensions

1. Introduction

Nestlé, the world's largest food company with 260.000 employees all around the world and 29 brands, has often faced extensive criticism concerning ecological and social exploitation. The allegations relate to immoral business with water, the destruction of the rainforest, pollution from plastic packaging and other issues of sustainability. Eckardt (2015), defines sustainability as "the establishment of sustainable and globally sustainable ways of living and doing business" (we translated, p. 5), which refers to economies that will not be detrimental to future generations. According to the Triple Bottom Line, sustainability includes the social component *People*, the ecological component *Planet* and the economic component *Profit* (Braccini & Margherita, 2019). All these three components have to be taken into account by companies in order to proceed in a fully sustainable way. The aspect of food sustainability is also relevant for the food company Nestlé. The *Food Life Cycle* involves the aspects of production, processing, packaging, ditribution, consumption and waste (Baldwin, 2015). All these aspects include sustainability components that must be considered for a sustainable economy (*ibid*.). This comprehensive number of aspects must not only be integrated into the company's machinations in form of measurements, they must also be communicated to the stakeholders in an appropriate manner in order to ensure the legitimation of the company. Sustainability communication serves this purpose. With its hundreds of millions of users worldwide and the ability to disseminate information in real time (Twitter Fundamentals, 2020), Twitter offers Nestlé the opportunity to reach a wide range of stakeholders, including new and existing customers worldwide as well as to address investors, partners and existing or new employees. Therefore, Twitter offers an optimal tool for the dissemination of sustainability communication among various relevant target groups and represents the communication platform of this study through which Nestlé communicates about sustainablity within different cultural contexts. According to Hofstede et al. (2010), national cultures can be clustered in different dimensions. The heart of every culture are the values that are gradually acquired and internalized from birth and form the starting point of every action (Hofstede et al., 2010). In their five cultural dimensions, namely power distance, uncertainty avoidance, masculinity vs. femininity, individualism vs. collectivism and long-term vs. short-term orientation, Germany is defined as less power-distant, more masculine, less uncertainty avoidant and more short-term oriented than France. Both countries don't differ a lot in the dimension of being highly individualist (Hofstede et al., 2010). Although the cultural dimensions have often been criticized of being generalising and essentialistic, they are listed among the most widespread cultural theories in international management, what made them relevant for this research.

2. Problem definition and Research Questions

As a global player, Nestlé is known worldwide for poor reputations regarding sustainability. To maintain its legitimacy towards international stakeholders, the company has to communicate sustainability measures and responsibility in a way that is appropriate for the target group, including the two major European markets Germany and France. According to the understanding of Hofstede et al. (2010), Nestlé must first become familiar with the peculiarities of cultures in order to report on sustainability in an appropriate and target-group-oriented manner. Indeed, the company has separate communication platforms on a country-by-country basis, indicating that communication differs from country to country and also between Germany and France. The question is, to what extent the differences in country-specific communication can be seen on the two Twitter profiles Nestlé Germany and Nestlé France and to what extent the cultural dimensions can be found in sustainability communication about ecological, social and economic topics and thus be used as an explanation for differences. Since France ranks first in the Food Sustainability Index 2019, and is ahead of Germany in this aspect, it can be assumed that especially

the communication about Food Sustainability differs between the countries because the importance is different. Against this background, the following research questions arose:

How does Nestlé's sustainability communication on Twitter differ in the German and French context?

→ What general differences can be identified in the sustainability communication between Nestlé Germany and Nestlé France?

→ To what extent can the country-specific differences in sustainability communication be determined using the cultural dimensions of Hofstede et al. (2010)?

→ What are the specific differences in communication about food sustainability between Nestlé France and Nestlé Germany?

3. Methodology

A comparative corpus analysis was used as a method. The corpus was defined by all publications on the German and French Nestlé twitter profiles, that have been published between 1st January 2019 and 31th March 2019. These tweets and retweets form the examination units. The investigation period was chosen since it was assumed that new communication approaches would also be implemented at the beginning of a new financial year. Twitter was defined as the investigation channel because of the characteristics of being a publication medium that is independent of time and place and which appeals to a wide variety of stakeholders. Additionally, a first overview showed that almost every publication on the both Nestlé Twitter profiles were about a sustainability issue. This fact made the medium particularly relevant as it proved its importance in terms of sustainability communication. The status of the tweets was the 18th February 2020. The measuring instrument, on the basis of which the counting of individual categories should take place, was developed by using a deductive-inductive method. In a first step, the data were examined for formal characteristics. Was it a tweet or a retweet? In case of retweets, from which profiles were they copied? Were quotes and reference used? Formal categories were formed from these characteristics. In a second step, the data were finally examined with regard to content and sustainability features and sorted into first content categories, which were further supplemented by multiple subsequent inductive-deductive recoding. The entire coding process was done by one single coder, Clarissa Glück. The content category determination was characterized by the application of the sustainability theory by searching for features of the *Triple Bottom Line*, the theory of food sustainability by searching for features of the *Food Life Cycle* as well as by own inductive category determination. A pretest was then used to test for validity and reliability, and to improve the selectivity of the categories. The final measuring instrument enabled examination on the macro level (what is the general weighting of the sustainability aspects *People*, *Planet*, *Profit*?), enabled on a micro level a detailed examination of individual sustainability issues and features of the communication concerning food sustainability. To consider the sustainability communication on the macro level, the three higher-ranking categories were deductively derived from theory of the *Triple Bottom Line*: *People* includes all tweets and retweets that address topics related to social welfare, education, justice, social resources, health and quality of life. *Planet* addresses all topics related to natural resources, water, biodiversity, the strike to keep oceans and environment clean or refer to sustainable agriculture. Finally, *Profit* includes all tweets and retweets that deal with economic sustainability, for example by addressing the training of future workers, recruiting processes, the maintenance of economic strength and the measures to increase it. Based on the above-mentioned sustainability categories, functional subcategories (level 3 categories)

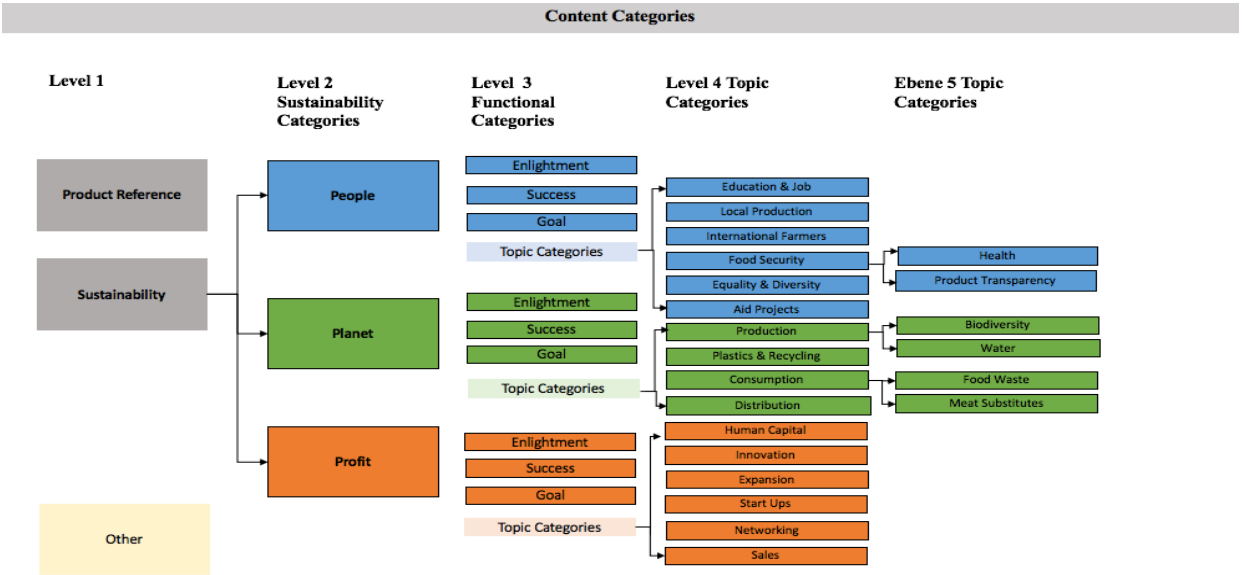


Figure 1: Content Categories

and topic categories (level 4 and 5 categories) where created afterwards. Some of these categories were identified inductively through the careful examination of the corpus, regardless from existing source material. Some other subordinated topic categories base deductively on mentioned sustainability categories and were taken over after a usability check on the corpus. One important content category that was created inductively is *Product reference*. This category provides information on whether the publication of a sustainability communication is directly or indirectly promoting a Nestlé product. It was included in order to examine more about the intention of the tweet. The category *Other* was built to include all remaining tweets without any sustainability content. After the identification of the usable categories, topic clusters were built and the category system created. The detailed hierarchical structure of the measuring instrument should make it possible to distance oneself from an essentialist view and to avoid a too superficial examination for differences (see Figure 1).

3.1. Formulation of Hypothesis

On the basis of the cultural dimensions (Hofstede et al., 2010), hypotheses were independently formulated, applying the characteristics of France and Germany to sustainability communication, which should be tested for falsification in the empirical part of the research. These hypotheses were thus used to answer the formulated research questions. The hypotheses are covered by self-defined indicators in the form of categories of the measuring instrument. It was assumed that, if Nestlé used the theory of cultural dimensions in the manner how sustainability is communicated on the French and German Twitter profiles, the following hypotheses would stand up to a falsification attempt and that the indicators would show significant differences between Nestlé France and Nestlé Germany:

- **H1 – dimension "power distance":** Nestlé France's sustainability communication is more focused, while Nestlé Germany takes diversity more into account. This is evident both in terms of the published sustainability issues and in formal terms.

A higher number of retweets and references indicate power distance. These are indicators of centralized communication, since Nestlé itself and some media as power entities would be the focus of communication.
In contrast to this, a more diverse use of sources, through the more frequent publication of retweets represent a low power distance, as this would indicate less centralized communication.

- **H2 – dimension "power distance":** Nestlé Germany's sustainability communication focuses more on social equality and diversity than in the case of Nestlé France.

The category *Equality & Diversity* serves as an indicator in this case. A significantly more frequent theming would therefore speak in favor of lower power distance.

- **H3 – dimension "uncertainty avoidance":** Nestlé France's sustainability communication shows a tendency to avoid uncertainty by avoiding communication about the unknown, while Nestlé Germany shows a tendency towards greater willingness to innovate through showing openness to new things.

Food safety, *health* and *product transparency* as well as *local production* are indicators of uncertainty avoidance, since they ensure transparency and address customer fears. On the other hand, indicators for a low level of uncertainty avoidance are *education*, as this provides information about existing problems and thus more likely to trigger uncertainties, as well as the categories *innovation*, *start-ups* and *expansion*, as these indicate openness for the unknown.

- **H4 – dimension "masculinity vs. femininity":** Nestlé France's sustainability communication is more based on empathy and compassion for the social and environmental circumstances, while Nestlé Germany focuses more on communicating long-term company growth.

Categories for masculinity are *profit*, *expansion*, *turnover*, *innovation* and *human capital*. These go hand in hand with economic growth. A more frequent *product reference* indicates a masculine orientation as well as this should stimulate consumption. On the other hand, femininity would show itself through a more extensive approach to topics related to solidarity and empathy. This includes the categories *planet* and *people in general*, *aid projects* and *international farmers*.

- **H5 – dimension "individualism":** Both, Nestlé France and Nestlé Germany highlight issues in social sustainability communication that affect the well-being of individuals, the customers. Collective topics are comparatively less addressed.

Health, *product transparency* and *food safety* indicate individualism, as well as *education & job*. The well-being of individuals is in the focus of communication. The category *international farmers* relates to the support of *small farmers* through fair production methods. *Aid projects*, refers to the support of people in need in developing countries. These are indicators of collectivism, since the wellbeing of society is in the foreground.

- **H6 – dimension "long-term orientation vs. short-term orientation":** Nestlé France's sustainability communications is increasingly geared towards long-term engagement, while Nestlé Germany places greater emphasis on speed and communicating solutions.

Categories for short-term orientation are *successes* and *goals* as they stand for quick solutions. A generally more frequent discussion of *profit* and *sales* and a higher number of *product references* would also be indicators for short-term orientation. In contrast, categories for long-term orientation are *planet*, which addresses long-term commitment, such as *biodiversity*, *water* and *plastic & recycling*.
Regarding the differences in the communication about food sustainability between the countries, following hypothesis was added:

- **H7:** Nestlé France reports more comprehensively on the aspects of food sustainability in comparison to other sustainability topics as well as in

comparison to Nestlé Germany's sustainability communication.

Categories for food sustainability are *local production*, *food safety*, production, *plastic & recycling*, *consumption* and *distribution*. These correspond to aspects of the Food Life Cycle, which defines food sustainability.

3.2. Performing the counting

The categories and thus the indicators for the hypotheses were always counted per tweet or retweet, i.e. always the entire publication, including the posting text, image or video and any links. Through the examination of Twitter content and thus computer-mediated communication it was not sure if the data could be called up indefinitely, so all the tweets were copied into a Word document. This ensured that the content was permanently available. Each examination unit was counted as often as categories could be found in it, i.e. separately for each identified category. In order to ensure the accuracy and selectivity of the categories and to be able to clearly assign the short messages, the coding had to be carried out several times and constantly refined. A clear definition of the categories and anchor examples made the assignment possible. These measures were necessary because of the high range of diverse content and thus indispensable for an effective counting. Since each examination unit had characteristics from several categories and could therefore be counted several times, percentage values for the coverage of each category were determined because the first examination was based on percentage coverage. When considering the percentage coverage, the hierarchy of the categories was taken into account in order to gain an insight into the thematic distribution at lower and higher levels. The differences and tendencies identified in these steps concerning the indicators of the hypotheses were checked for significance in the next step by making use of the Chi²-Test via the statistic program SPSS. A subsequent test for correlation for some of the categories by considering the Phi-coefficient should provide insights into the actual relationship between the indicators.

4. Results

First of all, it was striking that Nestlé France published 298 tweets which is more than twice as many research units in the relevant period on Twitter than at the Nestlé Germany profile. Here 132 Sustainability tweets were published. Since in both countries, almost all publications addressed at least one aspect of sustainability, one can speculate about a higher urgency with regard to sustainability communication in France. At the higher level, *People*, *Planet*, *Profit*, clear trends could be identified that differ between the countries. Nestlé Germany communicates more about economic aspects of sustainability, while Nestlé France reports more about nature and people. The subject of plastic in particular is of central relevance at Nestlé France's Twitter communication, whereas Nestlé Germany communicates a lot about supporting startups – a topic that doesn't even exist in the French communication. The following results can be summarized in relation to the hypotheses about the cultural dimensions: The results on the cultural dimension of power distance (Hofstede et al., 2010) show a tendency that is opposed to the assuming on the basis of the authors' country classification. Germany reports more centralized than France by using more tweets than retweets. Because Nestlé Germany quotes and references significantly more, sustainability communication is more diverse in this aspect, however, indicators about the origin of these references and citations could not be confirmed. Overall, Nestlé's cross-border sustainability communication is therefore not associated with any clear characteristics for different power distances. However, with regard to the cultural dimension of uncertainty avoidance, many indicators could be confirmed. Categories that were defined as indicators for openness to new and thus for low uncertainty avoidance were found significantly more within the communication of Nestlé Germany, while Nestlé France communicated significantly more about the categories that stand for uncertainty avoidance. Thus, the hypothesis that Nestlé France shows more signs of uncertainty avoidance and that Nestlé Germany shows greater willingness to take risks could largely be confirmed by the ways how they communicated sustainability.

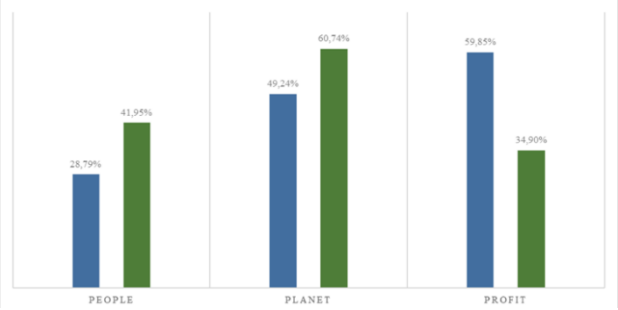


Figure 2: Thematic weighting of sustainability communication for France (Green) and Germany (Blue)

Concerning the dimension of masculinity vs. femininity, categories for economic sustainability and thus indicators for masculinity could actually be proven to appear more frequently at the communication of Nestlé Germany. Moreover, a product reference was used more often, which was also seen as an indicator of masculinity. At the same time, however, there is an ambivalence because Nestlé Germany also showed some indicators for femininity more frequently, such as the communication about aid projects and international farmers. Other indicators are more often addressed by Nestlé France. Therefore, the results can only identify some tendencies in terms of masculinity and femininity, whereas the hypothesis cannot be confirmed in general.

Regarding the dimension of individualism, the individualistic orientation of both countries is not reflected in sustainability communication, at least not with the help of the defined indicators. Although a few of the indicators for individualism are actually more common in the communication of both countries than indicators for collectivism, the other way around is also the case and

some of the indicators for collectivism have been counted more often. Also, no uniformity between the countries could be identified which, according to Hofstede et al., 2010, can both be assigned to individualism. Thus, the hypothesis cannot be confirmed.

Looking at the fifth cultural dimension, again only partial aspects of the indicators for both, long-term and short-term orientation, could be confirmed. Some others again indicated actually an opposite effect than expected by the hypothesis. While Nestlé France indeed tends to address more frequently environmental issues that involve a long-term commitment, the indicators that should confirm that Nestlé Germany highlights faster solutions by communicating successes and short-term goals, could not be confirmed.

The results of the Food Sustainability hypothesis partially reflect the results of how the two countries are listed in the Food Sustainability Index, however, it couldn't be confirmed that Nestlé France generally reports in more detail about all food sustainability indicators. Nestlé Germany communicates a lot about meat substitutes and thus especially highlights the consumption aspect. This reflects also the tendency towards a greater willingness to innovate and the stronger weighting of economic sustainability. In turn, Nestlé France's communication on food sustainability includes indicators that stand for uncertainty avoidance namely food security, consisting of the aspects *health* and *product transparency* as well as local production.

5. Conclusion

The results give insights into the sustainability communication of Nestlé via computer-mediated communication. As the results are based on tweets as one whole examination unit, future research should have a closer look on the usage of hashtags and single keywords. This would allow a more detailed insight in how sustainability communication is implemented via Twitter, whereas this research gives an overview of the thematic weighting. Twitter provides also huge potential by anchoring news through the usage of external links. Due to their size, these links could not be included in the present investigation, but they offer further information, while the tweets themselves, with their word limit, are only kept very concise. Of course, this conciseness also has an impact on the informative value, especially when applied on the complexity of culture, and thus also on the results. Future research should thus extend the focus on further information anchors, provided within the tweets. In addition, the huge difference in the number of published tweets between the two countries raises the question whether this could be due to different communication strategies in general. For example, the level of popularity and reach potential of Twitter in France and Germany could play a role. Therefore, it would be interesting if future research takes differences of general communication strategies of France and Germany into account, for example by examining the activation date of Twitter and the total amount of tweets, followers and interactions in both countries.

Moreover, the results show that just as culture is deeply rooted in a social group, it is also multi-layered, complex and dynamic and difficult to break down to five dimensions. Although tendencies towards differences between Nestlé France and Nestlé Germany can be recognized and the hypothesis of large and small uncertainty avoidance in sustainability communication was largely confirmed, the results of the other hypotheses do not lead to any clear generalizations about a possible classification of the two countries in the cultural dimensions. Nevertheless, the identified trends offer an approach for future research in this area. For example, Nestlé Germany seems to be communicating more about economic sustainability, integrating more product relevance and addressing innovation and corporate sustainability, while Nestlé France prefers social and environmental issues and places greater emphasis on food security and local origins. Absolute truths are not formulated, because various influences can play a role here. Nevertheless, the results could encourage a global player like Nestlé to think about a new direction towards networked, international sustainability management, which would certainly help to promote global awareness of global issues, to stimulate socially responsible behavior in relation to sustainability, and incidentally, to deepen the trust of the stakeholders with comprehensive, integrating sustainability communication.

6. References

Baldwin, C. (2015). The 10 principles of food industry sustainability. Wiley Blackwell. http://gbv.eblib.com/patron/FullRecord.aspx?p=403556 https://doi.org/10.1002/9781118447697

Braccini, A., & Margherita, E. (2019). Exploring Organizational Sustainability of Industry 4.0 under the Triple Bottom Line: The Case of a Manufacturing Company. Sustainability, 11(1), 36. https://doi.org/10.3390/su11010036

Ekardt, F. (2003). Theorie der Nachhaltigkeit: Ethische, rechtliche, politische und transformative Zugänge - am Beispiel von Klimawandel, Ressourcenknappheit und Welthandel (2. Auflage). Nomos Verlagsgesellschaft mbH & Co. KG. https://doi.org/10.5771/9783845254593

Hofstede, G [Geert], Hofstede, G. J., & Minkov, M. (2010). Cultures and organizations: Software of the mind ; intercultural cooperation and its importance for survival (Rev. and expanded 3. ed.). McGraw-Hill.

(2020, April 8). Twitter Grundlagen [Tweet]. Twitter. https://business.twitter.com/de/basics.html

Public Perception and Usage of Environmental Vocabulary: Building and Exploiting a Thematic Social Media Corpus

Tomara Gotkova, Nikolay Chepurnykh
ATILF, Université de Lorraine – CNRS
E-mail: tgotkova@atilf.fr, nchepurnykh@atilf.fr

Abstract

The main objective of our research is the study of core environmental vocabulary in English. Specifically, we examine the interplay between technical usage of this vocabulary – its scientific *terminology* – and its usage in ordinary discourse, with special attention paid to cases of terminological misuse. In the context of our study, by *ordinary discourse* we mean linguistic productions of general public found on social networks, more specifically, *Twitter* and *Reddit*. A key component of our study is data collection and data processing. In order to build our thematic environmental corpus using data from social networks, we compiled a set of environmental keywords using the term candidates extraction technique described in Shvets and Wanner (2020). Such a set was used to collect a test corpus of tweets and Reddit comments related to the environmental topic. The extracted raw texts were further processed to obtain a normalized and clean corpus. Additionally, in order to improve the quality of already processed data, a topic modelling technique was introduced to eliminate texts of no relevance to the environmental theme (Řehůřek and Sojka, 2010). Furthermore, using data from our initial corpus we built a subcorpus called *Carbon* to serve as a ground for a preliminary linguistic analysis of the environmental term *carbon*. In official environmental discourse, *carbon* is an "umbrella" term used to denote several things when it comes to greenhouse gas emissions. In addition, this term has settled in the ordinary laymen communication (Fletcher and Downing, 2011). By means of manual analysis complemented with automated techniques, we investigated how users of social networks conceptualize *carbon* and what linguistic means they employ to express those conceptualizations. Our findings indicate that the semantics of *carbon* is not unified in the official environmental terminology which allows for its ambiguous use and multiple interpretations in the specialized environmental discourse as well as in ordinary discourse.

Keywords: specialized vocabulary, social networks as corpora, the environment, linguistic analysis

References

Fletcher, J. and Downing, P. (2011). Consumer understanding of green terms: a supplementary report on consumer responses to specific green terms – *A report to the Department for Environment, Food and Rural Affairs*. Brook Lyndhyrst & Icaro Consulting, for Delfa, London.

Řehůřek, R. and Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May. ELRA.

Shvets, A. and Wanner, L. (2020). Concept extraction using pointer-generator networks and distant supervision for data augmentation. In C. Maria Keet et al., editors, *Knowledge Engineering and Knowledge Management*, pages 120–135, Cham. Springer International Publishing.

How Do Interlocutors in Instant Messaging Influence Each Other's Writing Style? Three Case Studies on Accommodation in Teenagers' Chat Conversations

Lisa Hilte, Reinhild Vandekerckhove, Walter Daelemans
University of Antwerp
E-mail: lisa.hilte@uantwerpen.be, reinhild.vandekerckhove@uantwerpen.be, walter.daelemans@uantwerpen.be

Abstract

We bring together three complementary case studies on the linguistic phenomenon of accommodation (i.e. people adapting their language use to that of their interlocutor) in Flemish teenagers' instant messages on Facebook Messenger and WhatsApp. Each case study focuses on a different aspect of the interlocutors' socio-demographic profile which may lead to distinct adaptation patterns: their gender, education, and age. Analyses with generalized linear mixed models reveal how teenagers' usage frequency for two sets of prototypical chatspeak features depends on the profile of the conversation partner. Significant adaptation patterns emerge with respect to all three socio-demographic variables. However, the linguistic features that are subject to adaptation and the accuracy of this so-called 'mirroring' differ depending on the interlocutors' profiles, which points to different accommodative behavior by distinct groups of teenagers.

Keywords: accommodation, mirroring, social media, teenagers, gender, education, age

1. Introduction

It has been attested repeatedly that teenagers' socio-demographic profiles and their online writing style correlate: youths with distinct profiles (in terms of e.g. gender or age) tend to favor certain markers of online writing to different extents (De Decker and Vandekerckhove, 2017; Hilte et al., 2020b; Verheijen, 2018). But little is known about the impact of the interlocutor's profile in online contexts such as instant messaging. While the phenomenon of accommodation (i.e. the adaptation of one's communicative behavior to that of one's interlocutor) has been widely investigated in spoken face-to-face interactions (see below), it is under-researched in (spontaneous, synchronous) written online conversations. The present paper aims to fill this gap by bringing together three complementary case studies on accommodation in teenagers' instant messages, with a respective focus on the interlocutors' gender, education, and age[1].

2. Related research

According to the sociolinguistic framework *Communication Accommodation Theory*, accommodation is driven by a desire to facilitate interaction and to regulate social distance between interlocutors (Dragojevic et al., 2015). Adapting one's language to that of others decreases the linguistic but also the social distance between interlocutors.

While the inclination to mirror the communicative style of others differs among individuals, some robust patterns relating to interlocutors' socio-demographic profiles have been found. With respect to gender, asymmetrical convergence has often been established, with women adapting their language use more strongly to men than vice versa (Palomares et al., 2016). But speech complementarity has been attested too, i.e. mutual divergence by men and women in order to consolidate social gender roles (Dragojevic et al., 2015).

In interactions between people of different ages/generations, common patterns are under-accommodation by older interlocutors versus over-accommodation by younger interlocutors, i.e. failing to adjust versus overadjusting one's communicative behavior to others (Giles and Gasiorek, 2011; Williams and Nussbaum, 2001). Overaccommodation to the elderly in particular is called 'patronizing talk', and includes e.g. oversimplified and excessively loud speech (Giles and Gasiorek, 2011; Williams and Nussbaum, 2001).

Interlocutors' educational background, finally, has - to the best of our knowledge - not yet been studied with respect to accommodation. Several papers do discuss the related variable of social power or position. So-called 'upward' social convergence has been attested repeatedly, that is more communicative adaptation towards interlocutors with greater (social) power (Dragojevic et al., 2015).

3. Data and method

3.1. Corpus

The corpus that is investigated in the three case studies contains 456,751 social media messages produced by 1,398 Flemish teenagers (living in Flanders, Dutch-speaking Belgium). The teenagers, aged 13-20, are pupils in secondary education. They attend general, technical, or vocational secondary education, which range from very theory- to mainly practice-oriented. The dataset contains the pupils' spontaneous, private instant messages, produced in Dutch on Facebook Messenger and

[1]Each of the case studies is published or forthcoming (Hilte et al., 2020a; Hilte et al., 2021; Hilte et al., under review). For an extensive discussion of the related work, research design and results per study, we refer to these papers. The present contribution brings these separate articles together for the first time and will focus on comparing and confronting the findings. This synthesis, in which studies on three distinct socio-demographic variables are confronted, offers a more complete and holistic perspective on accommodation.

WhatsApp, mainly between 2015-2016.

The corpus was collected in collaboration with schools. Pupils were invited to voluntarily donate chat conversations that were produced before our school visits. Note that some groups of pupils (e.g. girls) donated more data than others (e.g. boys). The participants also provided the relevant metadata: their own age, gender, and educational track (see Table 1 for the distributions in the corpus) and the age of their interlocutors (in case their data were not part of the teenage corpus). Finally, we asked the pupils' (and for minors also their parents') permission to store and linguistically analyze their anonymized messages.

Variable	Levels	Tokens
Gender	Girls	1,759,067 (66%)
	Boys	894,857 (34%)
Age	Young teens (13-16)	1,385,802 (52%)
	Older teens (17-20)	1,268,122 (48%)
Education	General	747,867 (28%)
	Technical	1,192,595 (45%)
	Vocational	713,462 (27%)
Medium	Facebook Messenger	2,045,396 (77%)
	WhatsApp	579,463 (22%)
Total		2,653,924

Table 1: Distributions in the corpus

3.2. Linguistic Variables

The case studies include two sets of linguistic variables that are prototypical of instant messaging, as they relate to two 'maxims' (implicit rules of linguistic conduct) of informal online writing, the maxims of expressive compensation and orality (Thurlow and Poff, 2013).

The principle of expressive compensation entails several (mostly typographic) strategies that compensate for the absence of certain expressive cues in written communication, such as facial expressions or voice volume. Examples of features are emoticons and emoji (e.g. :D), character repetition (e.g. *niiiice!!!*), and allcaps (e.g. *YES*).

The orality principle concerns speech-like writing: the register in instant messages often reflects typical speech patterns rather than classical written communication. In our corpus of Flemish teenagers' online writing, this results in the use of regional and colloquial features (e.g. *zoekt gij ambras?* for *zoek jij ruzie?*, 'are you picking a fight?'), and the insertion of English words or phrases typical of adolescent talk (e.g. *echt nice*, 'really nice').

All feature occurrences were detected automatically in the corpus with Python scripts. The software's performance was evaluated as reliable on a manually annotated test set (see Hilte et al. (2020b) for an extensive discussion).

3.3. Method

We approach accommodation from a quantitative perspective, i.e. as significant in- or decreases in the authors' usage frequency of linguistic features depending on their interlocutors' profiles. In all three case studies, generalized linear mixed models (Poisson distribution) were used to predict the participants' usage frequency for expressive and oral features, with aspects of their interlocutors' socio-demographic profiles serving as predictors. A random effect for subject and conversation was added to correct for repeated measurements, as both authors and conversations may occur more than once in the corpus (e.g. the same author in different interactions, or one interaction represented by its different participants). An observation-level random effect was added to avoid overdispersion (i.e. the variance of the response exceeding the mean; see Hilte et al. (2020b)). Finally, differences in sample size between observations are dealt with through an offset for (the logarithm of) the number of tokens per observation. Below, we discuss the best models per case study, i.e. the models that best fit the data (experimentally determined through stepwise deletion of insignificant predictors).

4. Results

Below, the results of the case studies are summarized. The present section and the discussion section aim to compare and confront the three separate analyses. For a more detailed discussion of the results per study (including e.g. model summary tables), see Hilte et al. (2020a; 2021; under review).

4.1. Gender

The first case study (Hilte et al., 2020a) compared teenagers' mixed-gender chats (including boys as well as girls; 34% of all conversations in the dataset) to same-gender chats (including only boys or only girls; 66% of all conversations in the dataset). The statistical analyses revealed no significant accommodation for oral markers (e.g. regional and colloquial language features), which are more prototypical 'male' features (Hilte et al., 2020b). But significant adaptation did emerge for expressive markers (e.g. emoji), which are generally inserted much more frequently by girls than boys (Hilte et al., 2020b). Both genders alter their use of expressive markers in mixed-gender chats, making their writing style more similar to that of their interlocutor of the opposite gender (see Figure 1). In one-on-one talks (including two interlocutors), which are shown on the left panel of Figure 1, girls insert significantly fewer ($p = 0.0266$) and boys significantly more expressive markers ($p < .0001$) when interacting with someone of the opposite gender versus when they chat with someone of the same gender. But, as Figure 1 shows, boys converge much more strongly to a (more expressive) 'female' style than vice versa. This contradicts previous work on spoken interactions, in which either a stronger convergence by women has been observed, or mutual divergence (see above). So the phenomenon of gender accommodation does extend from spoken to online written interactions, but the specific gender convergence patterns that have been attested in oral settings do not hold in our corpus of instant messages. Finally, we note that a similar

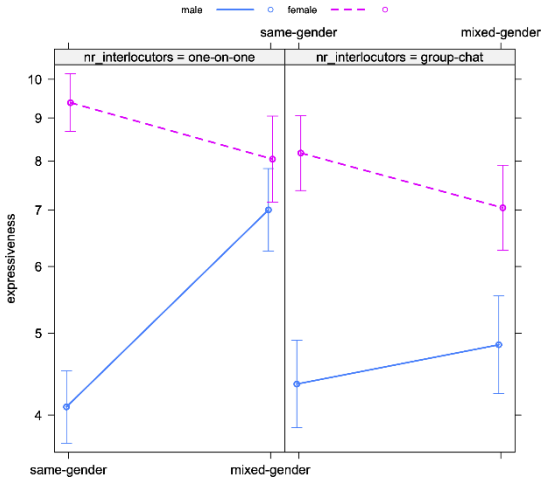


Figure 1: Expressive markers by author's gender, in same- vs mixed-gender conversations that are either one-on-one or group chats (predicted counts per 100 tokens).

girl1: we zullen er om 10 na 8 zijn 😉❤️💞 *We'll be there at 10 past 8*
girl2: Okeee ik zal zien dak klaarsta😉😍❤️ *Okay I'll make sure that I'm ready*
girl1: Toppp!! 😉😍❤️ *Great!*
girl1: Srry TUSSEN 8 en 10 na 😉😅❤️ *Sorry BETWEEN 8 and 10 past*
girl2: Oke ik zal klaarstaan❤️❤️❤️ *Okay I'll be ready*
girl1: Haha superr! 🏻👍🏻❤️❤️ *Haha super!*

Figure 2: Interaction between two female friends.

gender accommodation pattern as in one-on-one chats can be observed in group chats (including more than two interlocutors), as shown on the right panel of Figure 1. However, the accommodative adjustment by girls nor boys appeared significant (p = 0.07 resp. 0.1415). Intuitively, this finding of 'weaker' linguistic mirroring in group conversations makes sense – but we will come back to it in the discussion.

Since significant adaptation was only observed for expressive markers, which can serve as 'tools' for flirting, we examined to which extent gender accommodation can be obscured by flirting. An exploratory qualitative analysis of the corpus revealed that specific expressive markers with a romantic connotation (e.g. heart emoji) need not involve flirting, at least not in all-girls chats: girls use them abundantly in both romantic and non-romantic conversations (e.g. to express close friendship, as illustrated by the example in Figure 2).

Such a non-romantic use of these markers was absent in boys' chats. Consequently, we argue that flirting and gender accommodation (of expressive markers) are related yet distinct phenomena, and that even 'romantic' expressive markers are not solely used for flirting, but are truly part of a general female adolescent online style. Finally, the case study revealed that while the teenage girls in the corpus showed less accommodative behavior than the boys in terms of feature frequency (see above), they seemed to take into account male aversion for particular features (e.g. heart emoji) by avoiding them in non-flirty mixed-gender conversations. The boys, who accommodate more strongly in terms of feature frequency, much less do so in terms of adopting particular female features, except when flirting is involved.

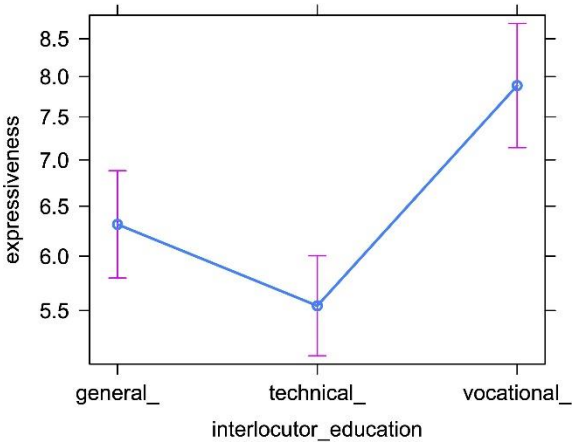


Figure 3: Expressive markers by the interlocutor's educational track (predicted counts per 100 tokens).

4.2. Education

The second case study (Hilte et al., under review) first of all shows that Flemish teenagers' instant messaging primarily proceeds within same-education networks, i.e. between pupils who attend the same educational track in secondary school (76% of the conversations in the dataset). But still, the statistical models reveal that when youths do chat across 'educational boundaries' (24% of the conversations in the dataset), they adapt their online writing style depending on their interlocutor's educational profile: the teenagers' usage frequency of oral and expressive markers significantly differs depending on whether they are interacting with a student from general, technical, or vocational education. However, this adaptation is only an accurate mirroring of the interlocutor's style for expressive features[2] (see Figure 3): i.e. the teenagers actually approach the average usage frequency for emoji, allcaps, ... by general, technical, and vocational pupils. For instance, when interacting with vocational students, who are generally the most ardent users of expressive markers (Hilte et al., 2020b), the teenagers tend to increase their own expressiveness to match that of their interlocutor. The frequency differences for expressive markers that are evoked by the interlocutor's educational profile (i.e. the teenagers' accommodative adjustments with respect to their conversation partner's education) are all significant (difference when chatting with a gen. vs tech. interlocutor p = 0.0082; gen. vs voc. interlocutor p = 0.0003; tech. vs voc. interlocutor p < .0001). The distinction between expressive and oral markers (for which respectively accurate versus inaccurate mirroring was observed) will be addressed in the discussion.

The findings of the second case study also indicate that pupils in different school systems do not adapt their online writing style to different extents (depending on their interlocutor's educational profile), so all teenage

[2]In the present contribution, we do not describe or visualize the pattern for oral markers, since it was not an accurate mirroring of the interlocutor's style (but see Hilte et al. (under review) for a discussion).

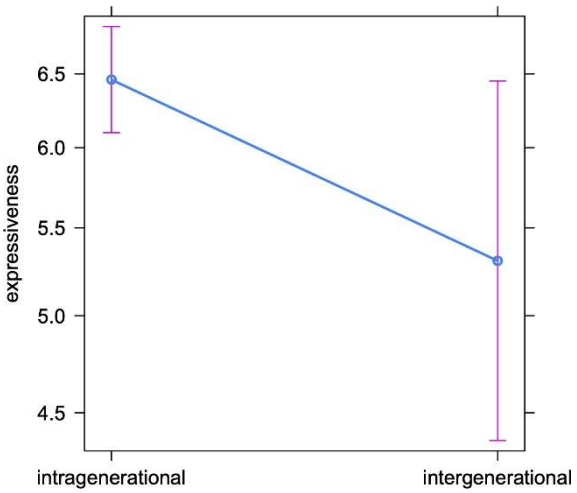


Figure 4: Expressive markers in intra- vs intergenerational conversations (predicted counts per 100 tokens).

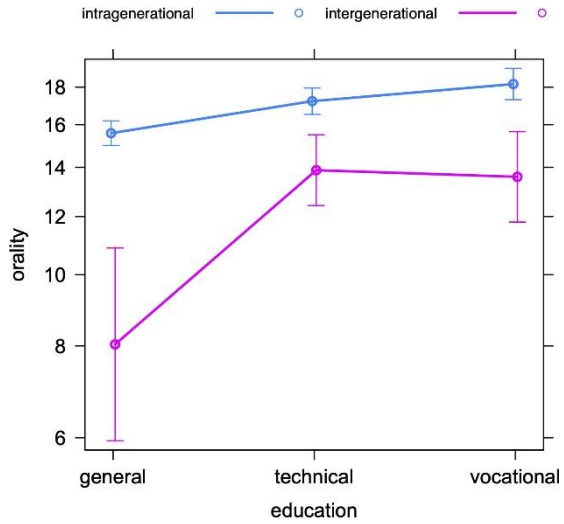


Figure 5: Oral markers by the author's educational track, in intra- vs intergenerational conversations (predicted counts per 100 tokens).

participants, regardless of whether they are general, technical, or vocational pupils, mirror their interlocutor's expressive writing to (more or less) the same extent. These symmetric accommodative efforts contrast with the asymmetric gender patterns that were observed in the first case study, with boys converging much more strongly to a 'female' style than vice versa. We will come back to this in the discussion.

Finally, the teenagers' adaptation to their interlocutors' educational profile is similar in one-on-one chats and group chats. That is quite surprising, since one might expect stronger mirroring in the former type of interactions for multiple reasons – see below.

4.3. Age

The third and final case study (Hilte et al., 2021) demonstrates how teenagers' instant messaging primarily proceeds within peer group networks (i.e. chats among teenagers only; 93% of the conversations in the dataset) and much less frequently across generations (i.e. chats including both teenagers and adults older than twenty; 7% of the conversations in the dataset). Still, the Poisson models show that when the teenagers do chat with older interlocutors, they adapt their own writing style by using significantly fewer expressive and oral markers (see Figures 4 and 5) ($p = 0.0473$ resp. $< .0001$). This decrease can be considered accurate mirroring, as (especially young) teenagers are more ardent users of these features than age groups above theirs (Hilte et al., 2020b; Prada et al., 2018; Verheijen, 2018). Consequently, youths appear to adapt their writing style to (that of) their older conversation partners by inserting fewer expressive markers (e.g. emoji) and fewer oral markers (e.g. colloquialisms) and thus adopting a more 'adult' writing style. This accommodation by the younger interlocutors in intergenerational communication is in line with previous findings on spoken interactions (see above). While the accommodation pattern for expressive compensation is not impacted by the teenagers' own socio-demographic profiles, the observed convergence with respect to oral markers differs depending on the teenagers' own educational track. A stronger adaptation – i.e. a stronger decrease in oral markers when chatting with older interlocutors – is made by pupils in the theory-oriented general secondary education (see Figure 5). So these pupils might have a stronger inclination to linguistic/stylistic mirroring (but further inspection is required). Such mirroring can be considered the product of meta-linguistic skills that are actually part of language teaching in more theory- oriented tracks only (VVKSO (2014)). In addition, a stronger command of written standard Dutch - which can be expected from theory-oriented students in view of their curriculum - might also increase these students' control of their standard versus speech-like rendition of Dutch words and phrases.

Finally, the teenagers' mirroring of older interlocutors' writing style is not any different in one-on-one chats compared to group chats. This echoes our findings for education but not for gender accommodation. The similarities and distinctions between the three case studies in this respect will be discussed below.

5. Discussion

This overview paper brought together three complementary case studies on linguistic accommodation in Flemish teenagers' instant messages, each focusing on a different aspect of the interlocutors' socio-demographic profiles. Convergence patterns emerged in all three studies: the teenagers adapt (aspects of) their online writing style depending on their conversation partner's gender, education, and age. However, differences could be noted concerning (a) the linguistic features subject to adaptation, (b) the accuracy of the mirroring, (c) its (a)symmetry, and (d) the impact of the number of interlocutors in a conversation. With respect to (a), in intergenerational interactions, teenagers converged to older interlocutors with respect to the use of both expressive and oral markers. In mixed-education talks, they adapted both their

expressive and speech-like writing too, though only the former adaptation actually mirrored their interlocutor's style (b). And in mixed-gender interactions, finally, significant convergence was attested for expressive markers only. A potential explanation for these differences concerns the higher 'visibility' of many of the (typographic/pictorial) expressive markers. The use of e.g. emoji may be more salient and therefore trigger more (accurate) convergence than the use of certain low-level colloquial markers. Furthermore, expressive features are generally used more deliberately, making them easier to manipulate, while teenagers (and especially pupils in practice-oriented tracks with a minor focus on formal Dutch writing) might be less in control over their speech-like rendering of Dutch words or the use of non-standard lexemes in general. It may thus simply be harder to (accurately) adapt speech-like writing.

With respect to the convergence pattern (c), symmetric accommodative efforts could be noted concerning interlocutors' educational tracks (i.e. pupils in different tracks did not adapt their online writing style to different extents), but not regarding interlocutors' gender: boys converged much more strongly to a 'female' writing style than vice versa. As for the age pattern, the frequency of oral markers in interactions with older interlocutors was suppressed more by pupils in theory-oriented tracks, which might again be linked to these pupils' stronger command of standard Dutch writing conventions.

We also compared one-on-one chats and group chats (d). Gender convergence was only significant in the former, i.e. instant messaging between one boy and one girl. The teenagers' adaptation towards interlocutors with different age or educational profiles, however, does not differ between these two conversational settings. That is quite fascinating, since stronger convergence might be expected in one-on-one interactions for several reasons. For instance, such interactions tend to be of a more intimate and personal nature than group chats, and trust is said to facilitate communicative convergence (Riordan et al., 2013). Second, linguistic mimicry is naturally more straightforward when there is only one other interlocutor to mirror. The distinction in this respect between the three case studies could indicate that gender accommodation mostly occurs in more intimate settings (while female vs male gender identities are expressed and emphasized more through gendered writing in group chats), in contrast to adaptation based on interlocutors' age or education accommodation.

Finally, recall that convergence narrows the linguistic and therefore also the social distance between interlocutors. The three case studies included in this paper suggest that the desire for social approval and closeness (a driving force behind accommodation) does not only hold among peers with similar socio-demographic profiles, but also across gender boundaries, educational boundaries, and across generations. In addition, note that mirroring in intereducational and intergenerational interactions is, in a sense, less obvious than in mixed-gender settings, as we observed how teenagers' instant messaging primarily proceeds in peer group networks within a same educational track. Still, both the educational background and the age of the interlocutor trigger accommodation.

6. Data availability statement

In order to protect the participants' privacy, and following the guidelines of our university's ethical committee, the collected dataset cannot be made publicly available. For more information on the database, see chapter 1 in Hilte (2019).

7. References

De Decker, B. and Vandekerckhove, R. (2017). Global features of online communication in local Flemish: Social and medium-related determinants. *Folia Linguistica*, 51(1):253–281.

Dragojevic, M., Gasiorek, J., and Giles, H. (2015). Communication accommodation theory. In C. R. Berger et al., editors, *The international encyclopedia of interpersonal communication (vol. 1)*, pages 1–21. Wiley Blackwell, London.

Giles, H. and Gasiorek, J. (2011). Intergenerational communication practices. In K. W. Schaie et al., editors, *The handbooks of aging consisting of three vols. Handbook of the psychology of aging*, pages 233–247. Elsevier Academic Press.

Hilte, L., Vandekerckhove, R., and Daelemans, W. (2020a). Linguistic accommodation in teenagers' social media writing: Convergence patterns in mixed-gender conversations. *Journal of Quantitative Linguistics*.

Hilte, L., Vandekerckhove, R., and Daelemans, W. (2020b). Modeling adolescents' online writing practices: The sociolectometry of non-standard writing on social media. *Zeitschrift fu¨r Dialektologie und Linguistik*, 87(2):173–201.

Hilte, L., Daelemans, W., and Vandekerckhove, R. (2021). Interlocutors' age impacts teenagers' online writing style: Accommodation in intra- and intergenerational online conversations. *Frontiers in Artificial Intelligence*, 4. Hilte, L., Daelemans, W., and Vandekerckhove, R. (under review). Communicating across educational boundaries: Accommodation patterns in adolescents' online interactions.

Hilte, L. (2019). *The social in social media writing: The impact of age, gender and social class indicators on adolescents' informal online writing practices*. Doctoral dissertation, University of Antwerp.

Palomares, N. A., Giles, H., Soliz, J., and Gallois, C. (2016). Intergroup accommodation, social categories, and identities. In H. Giles, editor, *Communication accommodation theory. Negotiating personal relationships and social identities across contexts*, 123–151. Cambridge University Press, Cambridge.

Prada, M., Rodrigues, D. L., Garrido, M. V., Lopes, D., Cavalheiro, B., and Gaspar, R. (2018). Motives, frequency and attitudes toward emoji and emoticon use. *Telematics and Informatics*, 35: 1925–1934.

Riordan, M. A., Markman, K. M., and Stewart, C. O. (2013). Communication accommodation in instant messaging: An examination of temporal convergence. *Journal of Language and Social Psychology*, 32(1): 84–95.

Thurlow, C. and Poff, M. (2013). Text messaging. In S.

Herring, et al., editors, *Pragmatics of computer-mediated communication*, 163–190. Mouton de Gruyter, Berlin and New York.

Verheijen, L. (2018). Orthographic principles in computer-mediated communication: The SUPER-functions of textisms and their interaction with age and medium. *Written Language & Literacy*, 21(1):111–145.

Vlaams Verbond van het Katholiek Secundair Onderwijs (VVKSO). (2014). *Nederlands derde graad aso-kso-tso. Leerplan secundair onderwijs.* VVKSO.

Williams, A. and Nussbaum, J. F. (2001). *Intergenerational communication across the life span*. Lawrence Erlbaum Associates, Mahwah and London.

Vitamin B: Bad or Benificial? A Social Network Analysis of Tweets on Perceived Risks in the Food and Consumer Product Safety Domain

Nina Lauran, Lidwien van de Wijngaert, Florian Kunneman
Radboud University Nijmegen
E-mail: n.lauran@let.ru, l.vandewijngaert@let.ru.nl, f.a.kunneman@vu.nl

Abstract

While our modern world provides more and more safety regulations, feelings of safety are declining. Hence, the relationship between safety and the perception of risk is paradoxal. The complexity of modern society combined with the development of a network society that communicates more and more horizontally urges to rethink risk communication practices. Prior research has used social media data to obtain insight into the way risks are communicated to audiences for single cases, while looking into the risk characteristics and communication development of multiple cases can further aid our understanding of the factors that trigger risk communication. In this exploratory paper, we zoom in on one of the cases, 'B6', using a combination of quantitative and qualitative methods to analyze the Twitter data about vitamin B6.

Keywords: risk assessment, social media, network analysis

1. Introduction

According to the Social Amplification of Risk Framework (SARF) risk not only requires technical assessment, but also initiates psychological, sociological, institutional and cultural processes (Kasperson et al., 1988). These processes can lead to an increased (amplification) or reduced (attenuation) perception of risk at the individual and at the societal level, and can influence the risk behavior of people.

In the SARF framework, the authors state that the perception of the severity of a risk is also determined by the way in which a risk is communicated. Experts themselves, but also media, influence how a risk is interpreted by lay people. Experts, institutions and media can act as an 'amplification station' (Kasperson et al., 1988). This framework is extensively used for the explanation of the amplification (exaggeration of risks) and attenuation (underreporting of risks) of risks reported by the traditional media. However, the rise of social media and its role in communicating and understanding risks has not yet been accounted for in the SARF framework (Lupton, 2016 in Fellenor et al, 2017). To gain further insights into evolving risk perception across groups on social media, we intend to address this gap by looking into the attenuation and amplification of risk in a corpus of Tweets about six different risk cases.

Shortly after SARF changed how scholars viewed risk through the transition from mass media to social media. In 1991, Giddens concluded that most citizens of industrialized countries live in a safer society than previous generations. Almost at the same time, Beck (1992) concluded that citizens are more concerned about risks in their daily lives than ever before. He classified the modern industrialized society as Risk Society in which negative stories in mass media prevail and risk and safety issues constantly dominate the policy and politics agendas. The establishment of governmental institutions and agencies at local, national and supranational scales did not reduce the public concern about risk as addressed by Beck. Leiss (1996) and Fischhoff (1999) concluded that a new perspective of risk communication by governmental organisations could be instrumental to close the gap between risks perceived by experts and by lay people, and to facilitate acceptance of small risks in favor of societal benefits. They, and many others, argued that engagement between risk assessors, risk managers, stakeholders and the general public is mandatory *towards a new modernity*.

At the time Giddens and Beck provided ground for the idea of a Risk Society, internet use took off, soon followed up by the development of a multiplicity of wireless communication tools and social media platforms such as Twitter. Castells (2010) argues that the shift from traditional mass media communication towards horizontal network communication clustered around digital platforms will drastically change modern society in general. He classified modernity as a Network Society.

Smith and colleagues (2021) propose that the ongoing social, technological and institutional transformations require that governmental organisations and agencies redesign their risk communication strategies and 'adapt its skills, working practices and structures to evolving audiences and their requirements'. This suggests that it is clear which audiences are targeted, and can be identified, and that the requirements of audiences are clear.

Lauran, Kunneman & Van de Wijngaert (2018) showed in their study of Twitter communication about Fipronil, that the way in which a risk is perceived and framed differs between groups and concluded that there is no such thing as one uniform audience.

This paper further investigates de development of Twitter audiences in non-incident/crisis situations and study an issue about food consumer product. With this study we aim to show how sub-arena social media audiences develop and how risks are under- or overreported (attenuated or amplified respectively) by the media. The results of the investigation help governmental institutions and agencies to redesign their future risk communication strategy, and can form a valuable addition to current insights about the amplification and attenuation of risk by looking into the role of Twitter in communicating and understanding risks.

1.1 Social media data

Social media platforms are a valuable data source for information about online communication strategies. Because social media data provide a continuous stream of real-time stored information, they open up possibilities for investigating the development of online opinion and networked relations with a study over an extended period of time (Lazer et al., 2009). Moreover, they enable the study of a risk or crisis situation and the feelings or thoughts involved without the need for respondents to look back on that event via a survey or interview. Social media data analysis allows the collection of a wide diversity of messages produced by a heterogeneous group of people directly at the time of the event (Gaspar et al., 2014). Searching the archived data for a specific timeframe provides the exact thoughts and opinions expressed by users on that particular day and it is possible to retrieve large amounts of data over a long period of time. The real-time information provided by social media platforms provides valuable insights into the development of a risk or crisis situation.

2. Method

The case is selected to obtain insights into the processes of attenuation and amplification. Design and Procedure
Tweets sent in Dutch and related to the subject that are posted between 1-1-2012 and 31-12-2020 are selected with use of the query 'vitamin B*', OR 'b6 OR b 6' and retrieved from Coosto[1]. Coosto is a social media monitoring tool with a focus on Dutch messages that scrapes the content of social media platforms, including Twitter, every 2 minutes and stores the data. A total of 2684 Tweets were retrieved for this B6-case. .
From the metadata of each Tweet in the corpus the date, author name, and text are used for further analysis. The date and name were not processed. The text column is preprocessed by removing interpunction and urls and setting all text to lowercase. Then all @mentions were taken to retrieve the user names of those mentioned or retweeted and to search for specific risk related words.

2.1 Quantitative analysis

The number of messages that were sent within the given timeframe, how the messaging developed over time, and the prominence and development of key words ('risk','danger','health','safety') are analyzed. In the corpus of the Tweets also the number of uses of these specific words over time is analyzed to compare how the risks are perceived in each case.

2.2 Network analysis

For each Tweet, the username of the sender is connected to the username of the one(s) mentioned or retweeted in the text of the Tweet to form an edgelist of all users that are connected to each other. Information about these relationships, such as the number of Tweets sent, the direction of the relationship and the hashtags and key words used are coupled in the edgelist as input for the network analysis.
The networks are visualized in the open-source tool Gephi (www.gephi.com) that provides measures and algorithms to visualize the imported edge list as a network. The usernames are used as input for the nodes and the connections between the users is used for the edges. Several algorithms are available for visualizing the network. In this research, we chose the ForceAtlas2 algorithm (Jacomy, Venturini, Heymann, & Bastian, 2014) as it is designed for networks which contain less than 10.000 nodes and therefore suits the characteristics of our data. This algorithm will allow visualization of the network.

- Density of the network is calculated by measuring the interconnectedness of the nodes.
- The PageRank of a node gives an indication of the importance of that node (the user) within the network, and a node is assumed to be more important if it has more links to other nodes (Brin & Page, 1998).
- The average degree is a measure for the amount of information that is passed from or to this user (node)
- Modularity calculates the number of different communities within the network.

These metrics are used in this research to look into distinctive groups within the network and to analyse how these groups differ in their communication about the risk event.

2.3 Qualitative analysis

We analyse users and messages to get more insight in the effect a message from that user can have and to analyze the communication pattern for each case. Therefore, we look into which users are central in the risk networks, what the content of the sent messages is, and what the most often retweeted messages are, since these examplify important Tweets.

3. Results

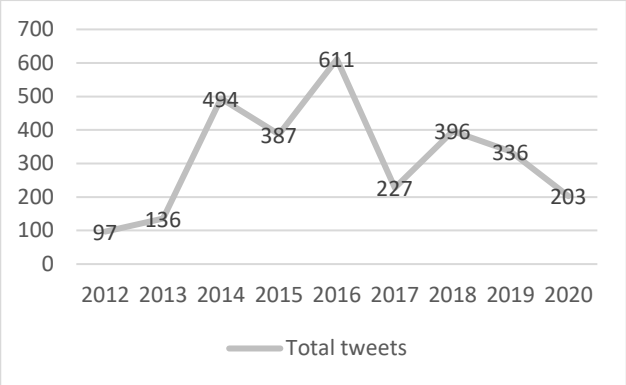
3.1 Case of Vitamin B6

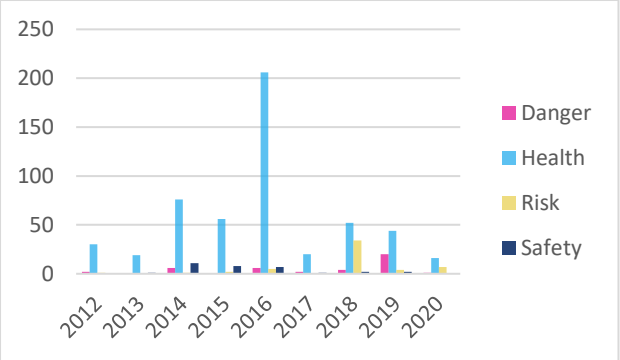
Vitamin B6 is an ambiguous risk: it is an important vitamin for metabolism and the immune and nerve system. But long-lasting daily use of highly concentrated B6 supplements can damage nerves. It is healthy and even important to consume some of it, but it can have adverse health effects when overusing it. For a long time, supplements of this vitamin were too concentrated. Nowadays, a limit has been set on the legal maximum dose a supplement may contain. In sum, it is a voluntary risk (people can decide whether or not to take the supplements), but the public was unfamiliar with it for a long time.

[1] www.coosto.com

3.2 Messages

A total of 2684 messages were retrieved with the query 'vitamin B6', 611(22,7%) of which in 2016.





In the graphs above, the total number of Tweets per year (above) and the times a Tweet mentioned one of the risk words (risk, safety, health, or danger (below)) is shown. Most messages were sent in 2016, the same year the most often mentioned term 'health' peaked. The number of Tweets strongly increased in 2014 and 2016 compared to the year before, and shows a strong decline in 2017 and after a small increase in 2018 slowly declines further. Over the whole period, the most often used term is 'health' and it's related words. Risk peaks in 2018, while danger is mentioned most often in 2019. The term safety is hardly used in this case.

The meaning of the word 'health' within the corpus of B6 Tweets changes over time. In 2012 and 2013, all but one of the uses of the word health are positive and relate to the health benefits of consuming B6. From 2014 onwards the balance between negative and positive associations of health with B6 shifts towards 'unhealthy'. In 2014 the exceedence of the limit becomes an important topic, leading to the negative connotation of adverse health effects of B6. In peak year 2016 the word 'health' is mostly used together with 'problems' (in 174 of 206 total 'health' mentions), since many people address the issue brought up in the most Retweeted message of that year.

3.3 Who and what

The five most retweeted messages provide insight in both the content of the messages and importance of the message within the case. These Tweets provide the most prominent type of user in this case as well: traditional media plays a significant role in the dissemination of information about the risk.

Message	# of RT	Year
"The advertisements for b12 supplements go through the roof. It becomes dangerous when people with a functional b12-deficit take these. They unknowingly blur their own diagnosis and keep getting sicker."	14	2019
Almost 1 in 5 users of vitamin B6 experience health problems"	15	2016
"*[Governmental organization]*: put a maximum on the amount of vitamin b6 in food supplements"	20	2017
"Watch out for quacks! – leg injury *[Dutch athlete]* is to blame to too much vitamin b6"	21	2015
From today the commodity law states that supplements cannot contain more than 21 milligram vitamin b6. Those who get in too much b6 for a long time are at risk of the nerve condition periferal neuropathy"	32	2018

Table 2: The 5 most retweeted B6 messages.

What strikes when looking at the retweets in chronological order, is that a pattern emerges. In 2015, the link between vitamin B6 and adverse health effects became public through the injury of a Dutch speed skater. In 2016, a Dutch television show addressed the risk of consuming B6 supplements, and in 2017, the Netherlands Food Safety Authority advised to lower the maximum amount of B6 a supplement may contain, but it took until 2018 to solidify this advise in law. Four years have past between the first signs of adversary health effects and actual change in laws lowering the maximum amount. After that, in 2019, the risk focus changes towards another type of vitamin; B12.

4. Network

In the network visualization of the Tweets, the different groups are colored in greyscale and the key words mentioned by the users in red for danger, light blue for health, dark blue for safety and yellow for risk. This network depiction shows that different groups of users differ in their use of 'risk-words', as some use none, while other groups use 1 term and a few times more than 1 term is used.

Figure 1: Network of B6 tweets

5. Discussion

For the B6 case, we saw that the development of the number of Tweets over time had a clear upward trend in 2014 and 2016, with most Tweets sent in 2016. These two peaks can be explained by the attention the subject got from Dutch television programmes, where several people shared their experiences with the adverse effects of overusing the vitamin and consumer protection organisations alarming government about the risks. The Tweets sent by and about the television programme in 2016 were about how common these adverse health effects were, thus often using the word 'health'. Although the issue was already addressed in 2014, it took until after the broad attention it got the second time for the government to take action by announcing that the maximum amount of B6 a supplement may contain will be lowered, and law was adjusted. When word spread that this issue was no longer a risk since overusing the vitamin was legally contained, the discourse changed and the discussion went on about other types of vitamins.

When zooming in on the use of the key terms associated with risk, a notable change in the use of the word 'health' is visible. At first, the vitamin B6 was associated with healthy living an eating. However, together with the first signal of adverse health effects by a Dutch television programme, the association of the word health changed from positive to negative. The risk was already on the radar of traditional media, which became visible in the Twitter traffic. While the media reported the risk frequently, it seems like the tipping point of attention was reached after the second wave of media coverage, rippling towards the political arena and policy change. When the new law was enforced, attention for this specific vitamin deminished and refocused on other comparable risk events. This case shows that traditional and social media play an important role in creating public awareness of the risk and by sharing personal stories of adversary health effects. This attention for the risk lead to policy change in this case.

In this research, we combined content analysis with network analysis to provide insight into the mechanisms of online communication about risk. This combination aids the understanding and interpretation of online risk communication and the social amplification of risk. This analysis makes the development of the risk and phases of attenuation and amplification more visible, because the content of the discussion within subgroups can be analysed, thus generating detailed insights into the stakeholders and their views. The gain of using social media data for SARF is that it not only provides information from traditional media, but also depicts the direct linked reactions of the public to that information. Furthermore, all parties that have a stake in the risk communication about the event can share their opinions directly and communicate horizontally with others. The detection of the use of key words over time and in relation to other users teaches us more about how risks are defined and how different audiences communicate about the risk. This knowledge is valuable for further developing the SARF-framework and to understand how different audiences react and post messages about risk.

The risk case selected in this research is Dutch, and although these risks are present around the world, it is not to say that every audience will react in the same way as it did here. Future research could focus on other cases and a comparison between countries in order to establish common ground for the specific characteristics of the risk event and the reactions it provokes in the public. For the Dutch situation we continued our research with 5 other non-crisis cases (data not presented).

6. Conclusion

The combination of methods improves our understanding of online risk communication. As discussed in the introduction, too little attention has been paid to social media and its role in communicating and understanding risk in SARF. This study shows that social media plays an important role in the communication about risks, both via messages shared by individual users as by further spreading the messages sent by traditional media. Furthermore, this research confirms Castells' network society vision in that social media, in this case Twitter, opens up possibilities for direct horizontal communication between media, individuals, NGO's, stakeholders, and government. This research aids in identifying audiences within the online communication about risks and forms a starting point for further research into the specific targeting of audiences.

7. References

Beck, U.(1992). From industrial society to the risk society: Questions of survival, social structure and ecological enlightenment. *Theory, Culture & Society, 9*(1), 97-123.

Brin, S., & Page, L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks, 30,* 107–117.

Castells, M. (2010) *The rise of the network society.* Chichester: Wiley-Blackwell.

Fellenor, J., Barnett, J., Potter, C., Urquhart, J., Mumford, J.D., & Quine, C.P. (2017). The social amplification of risk on Twitter: the case of ash dieback disease in the United Kingdom. *Journal of Risk Research*.

Fischhof (1999). Why (Cancer) Risk Communication Can Be Hard. *JNCI Monographs, 1999*(25), 7–13

Gaspar, R., Gorjão, S., Seibt, B., Lima, L., Barnett, J., Moss, A, & Wills, J. (2014). Tweeting during food crises: A psychosocial analysis of threat coping expressions in Spain, during the 2011 European EHEC outbreak. *International Journal of Human-Computer Studies, 72,* 239-254.

Giddens, A.(1991). *Modernity and Self-Identity. Self and Society in the Late Modern Age.* Stanford: Stanford University Press.

Jacomy, M., Venturini,T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLOS ONE, 9,* e98679.

Kasperson, R.E., Renn, O., Slovic, P., Brown, H.S., Emel, J., Goble, R., Kasperson, J.X. & Ratick, S. (1988). The Social Amplification of Risk: A Conceptual Framework. *Risk Analysis, 8*(2).

Lauran, N., Kunneman, F., & Van de Wijngaert, L.(2020). Connecting social media data and crisis communication theory: a case study on the chicken and the egg. *Journal of Risk Research, 23*(10), 1259-1277.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, AL., Brewer, D., Christakis, N., et al. (2009). Social science. Computational social science. *Science, 323*(5915), 721–723.

Leiss, W.(1996). Three Phases in the Evolution of Risk Communication Practice. *The annals of the AA Political and Social Science, 545*(1), 85-94.

Smith, A., Vrbos, D., Alabiso, J., Healy, A., Ramsay, J., & Gallani, B. (2021). Future directions for risk communications at EFSA. *EFSA Journal, 19*(2).

Webcare Across Public and Private Social Media Channels: How Stakeholders and the Netherlands Red Cross Adapt Their Messages to Channel Affordances

Christine Liebrecht, Charlotte van Hooijdonk
Tilburg University, Utrecht University
E-mail: c.c.liebrecht@tilburguniversity.edu, c.m.j.vanhooijdonk@uu.nl

Abstract

Previous research has focused on the characteristics of webcare conversations held on public social media channels, such as Twitter and Facebook (e.g., Huibers & Verhoeven, 2014; van Os, Hachmang & van der Pool, 2016). However, webcare conversations are increasingly held on private social media channels, such as Twitter DM and Facebook Messenger (e.g., van Os et al., 2018). It is still unknown to what extent stakeholders and organizations adapt their messages to the affordances of public and private social media channels.
This paper reports a content analysis of webcare conversations (n = 423) between stakeholders and the Netherlands Red Cross on public (Twitter, Facebook, Instagram) and private (Twitter DM, Facebook Messenger, WhatsApp) social media channels. By annotating the stakeholders' messages on their motives (i.e., question, remark, complaint, or compliment) for contacting the organization at the beginning of the conversation, and the sentiment of the last stakeholder's message, we investigated to what extent their motives match the channel affordances indicated in prior research. We also annotated the webcare messages on their communication style (i.e., Conversational Human Voice, CHV; Kelleher, 2009), to examine whether webcare practitioners differ in the adoption of CHV across different social media channels. CHV was annotated using the identification instrument of van Hooijdonk and Liebrecht (2018) who differentiated several linguistic elements of CHV belonging to three main tactics distinguished in literature: message personalization, informal speech, and invitational rhetoric (van Noort, Willemsen, Kerkhof & Verhoeven, 2015).
The findings reveal private channels mainly serve the purpose of customer service: stakeholders approach the organization with questions, and the organization uses message personalization to enhance the experience of one-to-one communication. Public channels mainly serve the purpose of reputation management: stakeholders post remarks and compliments and the organization adapts the communication style of its messages to the affordances of the individual platform.

Keywords: webcare, conversational human voice, the Netherlands Red Cross, public social media, private channels

References

Kelleher, T. (2009). Conversational voice, communicated commitment, and public relations outcomes in interactive online communication. *Journal of Communication, 59*(1), 172-188.

Huibers, J., & Verhoeven, J. (2014). Webcare als online reputatiemanagement: Het gebruik van webcarestrategieën en conversational human voice in Nederland, en de effecten hiervan op de corporate reputatie. *Tijdschrift voor Communicatiewetenschap, 42*(2), 165-189.

Van Hooijdonk, C., & Liebrecht, C. (2018). Wat vervelend dat de fiets niet is opgeruimd! Heb je een zaaknummer voor mij? ^EK. *Tijdschrift voor Taalbeheersing*, *40*(1), 45-81.

Van Noort, G., Willemsen, L. M., Kerkhof, P., & Verhoeven, J. W. (2015). Webcare as an integrative tool for customer care, reputation management, and online marketing: A literature review. *Integrated communications in the postmodern era*, 77-99.

Van Os, R., Hachmang, D., & van der Pool, E. (2016). Webcare-strategieën door OV-aanbieders. *Tijdschrift voor Communicatiewetenschap*, *44*(3).

Van Os, R., Hachmang, D., Akpinar, M., Keuning, A. & Derksen, M. (2018). *Stand van Webcare 2018*. Consulted on 11 June 2021, via: https://www.upstream.nl/wpcontent/uploads/2018/09/20180918-Onderzoek-Standvan-Webcare-2018.pdf.

Political Internet Memes and Digital Activism on Facebook: A Multimodal Corpus-Based Analysis of China's Diba Expedition to Hong Kong

Zhe Liu
University of Leeds
E-mail: mlzl@leeds.ac.uk

Abstract

As a representative practice of vernacular creativity (Burgess, 2006), Internet memes often facilitate public commentary and online communication (Milner, 2013). Within the digital environment, Internet memes not only function as mirrors of cultures and societies, but also present creators' skills of deploying different semiotic resources to make meaning.

In 2019, a group of Chinese young netizen bypassed China's Great Firewall, a nationwide network security system. Then, they collectively posted pro-government Internet memes as comments on Facebook pages of several Hong Kong institutions and public figures. These online activities are known as the Diba Expedition, which aims to express dissatisfaction to the targeted public institutions and figures on specific political issues. The self-generated Internet memes that were transmitted in each expedition are mainly comprised of images and writings.

Although the Diba Expedition has been previously investigated in terms of mobilisation and organisation (Yang et al, 2017; Liu, 2018), little attention has been placed on the practices of meaning-making and the engagement in digital activism through multimodal Internet memes.

This study focuses on 30 Internet memes, which were collected from the Facebook page of Joshua Wong, one of the Hong Kong pro-democracy activists. Using the GeM (Genre and Multimodality) model (Bateman, 2014) and a modified annotation schema, a corpus is built to facilitate a social semiotic multimodal analysis (Kress, 2010) in terms of (1) the functions that images and writings in the data respectively have during the process of making meaning, and (2) the intersemiotic relations between them.

Findings shows that the multimodal components of the analysed data are not only purposefully combined for expressing desired semantic meanings, but also creatively borrowed from other contexts for communicating with potential viewers of different backgrounds.

Keywords: Internet memes, multimodality, social semiotics, corpora, China

References:

Bateman, J.A. (2014). Developing a GeM (Genre and Multimodality) model. In Norris, S. and Maier, C.D. (eds) Interactions, Images and Texts: A Reader in Multimodality. Berlin: De Gruyter, 25-36.

Burgess, J. (2006). Hearing ordinary voices: cultural studies, vernacular creativity and digital storytelling. *Journal of Media & Cultural Studies*, 20(2), 201-214.

Kress, G. (2010). Multimodality: A Social Semiotic Approach to Communication. London: Routledge Falmer.

Liu, H. (2018). *From Cyber-nationalism to Fandom Nationalism: The Case of the Diba Expedition in China*. London: Routledge.

Milner, R.M. (2013). Pop polyvocality: internet memes, public participation, and the Occupy Wall Street movement. *International Journal of Communication*, 7, 2357-2390.

Yang, S., Chen, P.Y., Shih, P.C., Bardzell, J., and Bardzell, S.W. (2017). Cross-strait frenemies: Chinese netizens VPN in to Facebook Taiwan. *PACM on Human-Computer Interaction*, 1, 1-19.

Detecting Facial Emotions to Support Linguistic Analysis of Political Tweets: a Multimodal Approach

Julien Longhi
CY Cergy Paris université, AGORA, IDHN, IUF
E-mail: julien.longhi@cyu.fr

Abstract

This paper focuses on the analysis of audiovisual content embedded in tweets, especially in a political context. It is based on research conducted since 2013 on digital discourse on Twitter, and addresses a new challenge for the CMC community: the automatic consideration of multimedia content embedded in messages. To this end, we propose a promising research direction: the automatic extraction of emotions from political personalities, using artificial intelligence. Although this work is still in its early stages, it offers promising possibilities for discourse analysis, while requiring some precautions. It also allows us to consider new solutions for the linguistic analysis of emotions or feelings, by providing additional labels for annotating corpora.

Keywords: political tweet, sentiment analysis, emotions, deep learning

1. Introduction

Social media are increasingly integrating multimedia content. This is particularly the case for tweets which in their body contain, in addition to text, visual or audiovisual elements (Simon et al., 2016) that contribute to the argumentation put forward. For example, in the following tweet, an extract from a speech by presidential candidate Emmanuel Macron is quoted by the user, but the quote is linked to its broader context in the form of a video:

Figure 1: Capture of a tweet integrating a video of Emmanuel Macron

In this paper, we extend our work conducted since 2013 on digital discourse on Twitter (Longhi 2013; Djemili et al. 2014; Longhi 2020) and address a new challenge for CMC corpora research: the automated consideration of multimedia content embedded in social media messages. To this end, we propose a promising research area whose scientific progress is remarkable: the automatic extraction of emotions from political figures, based on tools using artificial intelligence. This exploratory paper is structured as follows: first, we delimit the main lines of the research by situating ourselves in relation to different approaches; second, we briefly present some criteria for describing faces, using an example from Emmanuel Macron's video; third, we look at their use for identifying emotions and propose a multimodal analysis of a video sequence by bringing together several observations; fourthly, we add to these results a linguistic analysis based on BERT, a deeply bidirectional, unsupervised language representation. We then conclude on the significance of such a study (which extracts meaningful elements from videos that are useful for linguistic analysis) and present a way to encode emotions in corpora using the TEI format.

2. Delimitation of the study area

According to Scherer et al. (2011: 426), "emotion expression and impression are determined by an interaction of psychobiological and sociocultural factors", and "the question of whether [the] expression and impression of emotion are either universal or culturally specific in terms of a dichotomy is therefore moot, as science has progressed to a higher level of understanding". Of course, cross-cultural issues related to the automatic detection of emotion need to be addressed, and ontological positions related to the labelling of multimodal and linguistic content from the outside and from the participants' point of view which we will integrate also need to be explored.

To do so, we plan to develop the etic and emic perspectives in investigating how emotions are expressed and interpreted by others. Indeed, the emic/etic concepts, coined by Pike, are "two basic standpoints from which a human observer can describe human behavior, each of them valuable for certain specific purposes" (Pike, 1954: 8). Mostowlansky & Rota (2020) point out that, according to Pike, "an etic approach would rely on a generalised classification system devised by the researcher in advance for the study of any particular culture in order to compare and classify behavioural data from across the world", while "an emic approach would dispense with a priori means of classification". For the moment, our research has been focusing on one culture at a time, but for future work we intend to integrate an emic approach (which would "call attention to the fact that two etically identical behaviours can in fact differ profoundly, depending on the meaning and purpose of the actors").

3. Recognition of action units

Soysal et al. (2017) propose "a new deep learning scheme that integrates convolutional neural networks with association rule mining", according to the "action units" (AU) described on faces:

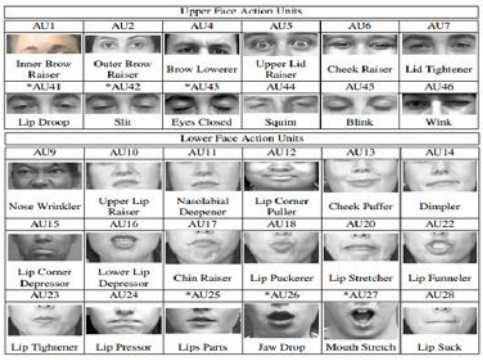


Figure 2: Action units (Saysal et al., 2017)

We can see above an inventory of different facial expressions, for both the upper and lower face, which we can apply to our video as below:

Figure 3: Eyebrow (left), Mouth (right)

4. Video analysis

To do an automatic analysis of this video, we used OpenFace[1] which provides "free and open source face recognition with deep neural networks". We applied this tool to a much commented "punchline", *Madame Le Pen, la France mérite mieux que vous* ("Madame Le Pen, France deserves better than you"), lasting 3 seconds. When analysing the video, OpenFace (Amos et al., 2016) proposes a face model with a large number of points:

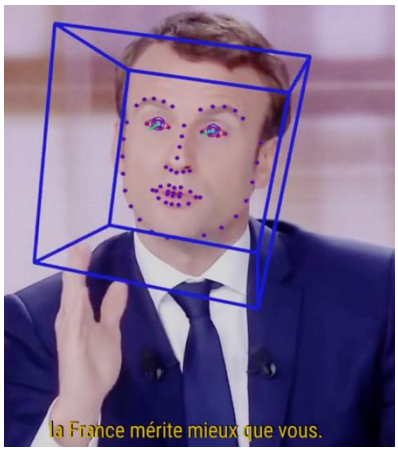


Figure 4: Face modelling

The video is therefore decomposed into a set of images, and each is automatically analysed according to the AU described in Figure 2.

4.1. Segmentation and interpretation of results

Once this division is made, we can observe the different AUs on graphs:

Figure 5: Presence of AU according to the timestamp of a sequence

Thus, we can observe the "strength" of each AU by "frame", which then allows us to make assumptions about potential facial emotions. Indeed, as some work shows, the combination of different AUs potentially allows us to define some basic emotions, as illustrated for example by Wegrzyn et al. (2017):

[1] https://cmusatyalab.github.io/openface/. AU extraction with OpenFace was performed with Aymeric Erades, research engineer at IDHN in 2021.

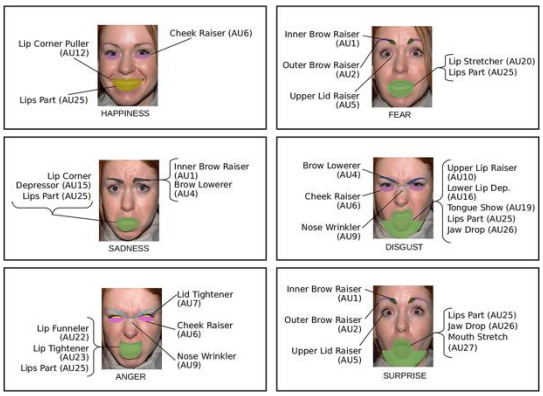


Figure 6: Definition of six emotions by Wegrzyn et al. (2017)

The hypothesis we make is that we could detect emotions in order to add this information to the analysis of the messages posted. To do so, and in an ecological approach to corpora, it is necessary to be able to return to the image corpus and verify the automatic results.

4.2. Methodological precaution: categories of emotions and using the corpus

According to Crawford (2021), there are several limitations to emotion detection:

- The emotion categories themselves are problematic: they are often based on Ekman's pioneering work and "we should question the origins of those categories themselves, as well as their social and political consequences". For us, therefore, this work of detecting emotions comes in "support" of linguistic analysis and is not an end in itself;
- As far as detection is concerned, "Emotions are complicated, and they develop and change in relation to our cultures and histories": it is therefore important to be able to return to the corpus to observe the data and validate the results.

For this purpose, OpenFace allows us to return to the identified frames and related images:

Figure 7: Back to the video corpus

We can thus prevent potential mistakes being made by the system and observe precisely some key moments whose interpretation would seem to be relevant.

5. For a multimodal analysis of political tweets

This work allows us to partially overcome the difficulty related to the multimodal dimension of the messages written on Twitter. Indeed, in the case of the quote that interests us here, Macron's account limits its sharing to the textual content only:

Figure 8: Tweet of Emmanuel Macron

This message is interesting from a linguistic point of view: to quantify it, we applied two models developed with BERT to capture the intensity or quality of the sentiments expressed. We chose BERT because, according to Tenney et al. (2019), "deep language models can represent the types of syntactic and semantic abstractions traditionally believed necessary for language processing, and moreover that they can model complex interactions between different levels of hierarchical information". In particular, "BERT is the first deeply bidirectional, unsupervised language representation, pre-trained using only a plain text corpus".[2] Even if sentiment analysis is a burning issue, Essebar et al. (2021) demonstrate that "French PTM shows improvements of text representation in many NLP tasks including sentiment analysis at sentence-level".

We thus applied two pre-trained models based on BERT and when we looked at our sentence we found that it rather expressed anger, with strong intensity:

[2] https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html

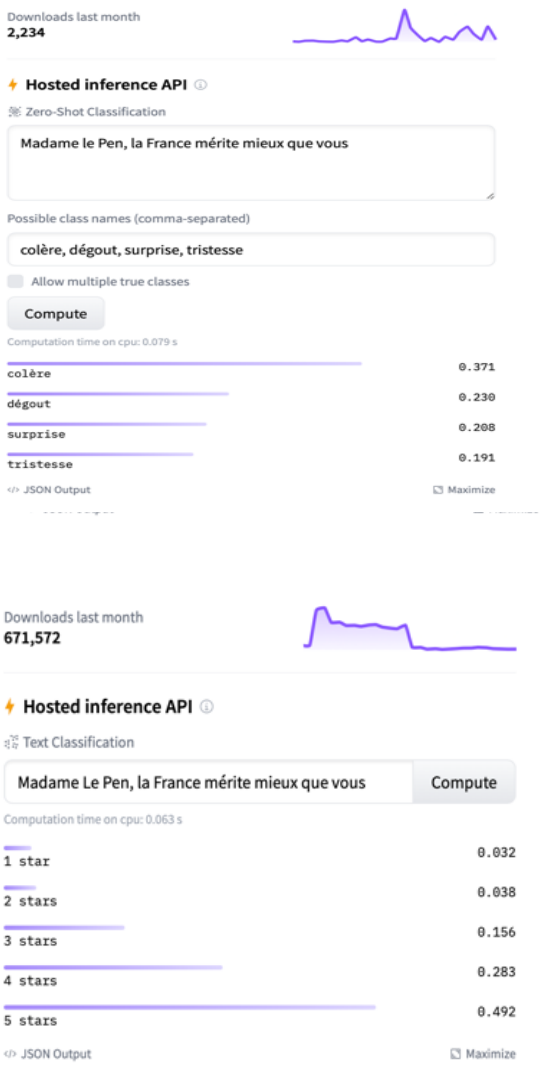


Figure 9: Linguistic analyses based on BERT models

The sentence is identified as anger (with a score of 0.371) with an intensity of "5 stars" (0.492). But since, as we have seen, this statement was part of an interaction, we can try to know if the candidate also physically expressed this anger. In figure 5, we proposed a graph, called AU50_r, which represents the detection of anger (by combining AUs 6, 7, 9, 22, 23, 25):

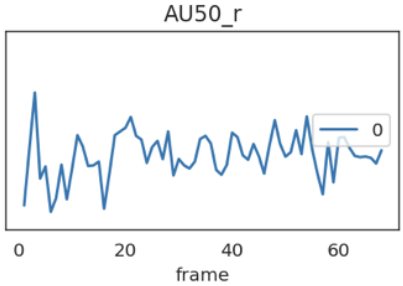


Figure 10: Representation of the emotion "anger" in the short extract

We can now confirm that the automatic analysis of the candidate's face reflected a certain anger. This result met our initial objective of introducing a multimodal dimension to the automated analysis of political messages on Twitter. In addition to enriching the corpus, this solution for the linguistic analysis of emotions or feelings (by providing additional labels for annotating corpora) could be a precise help for automatic language processing by making the annotation of corpora reliable, which could then be used to train models with the aim of improving emotion or feeling recognition algorithms.

6. Conclusion

This paper, which opens up new perspectives of research for the linguistic analysis of political tweets, allows us to obtain information from the multimedia content embedded in tweets, in order to go beyond the limitations of a sole textual analysis. We can imagine, in the short term, that these analyses could be integrated as annotations in corpora, with a set of basic emotion labels or general descriptors. We could thus have, in the same way as temporal or material variables, an "emotion" variable using an xml tag, for example, in the TEI representation of tweets:

<post xml:id="cmr-présidentielle-a859882207648854016" who="#cmr-présidentielle-p1976143068"
when="2017-05-03T23:27:48.0" xml:lang="fra">

Madame Le Pen, la France mérite mieux que vous. <distinct type="twitter-hashtag"><ident>#</ident><rs ref="https://twitter.com/search?q=%232017LeDébat&src=hash">2017LeDébat</rs></distinct>

<trailer>
<fs>
<f name="medium">
<string>Twitter Web Client</string>
</f>
<f name="favoritecount">
<numeric value="9828"/>
</f>
<f name="retweetcount">
<numeric value="9625"/>
</f>
<f name="emotion">
<emotion name="anger"/>
</f>
</fs>
</trailer>

This would allow, at first, to take into account the facial emotion expressed when such or such a sentence is uttered. Of course, this opens the way to many other aspects (gestures, mimics) on which we will work in future research.

More generally as far as the analysis of political discourse is concerned, this perspective allows for a more complete grasp of political argumentation since, according to Lara et al. (2016: 155), "emotion is not only a dimension working

as an enhancer or mitigator for the persuasive aim", "it is used as an argument itself". This allows us to better understand political discourse, especially on Twitter, by not neglecting an important part of the contribution made by the staging of emotions. To improve that, future work should include intercultural specificities because, "since emotion recognition is culturally dependent, it is predicted that a culturally inclusive recognition model should yield better results than a model that does not consider culture" (Quiros-Ramirez & Onisawa, 2015). This is another challenge to be taken up in order to offer a better match between the analysis model and possible application contexts.

7. References

Amos, B., Ludwiczuk, B., Satyanarayanan, M. (2016). Openface: A general-purpose face recognition library with mobile applications. CMU-CS-16-118, CMU School of Computer Science, Tech. Rep.

Crawford, K. (2021). Artificial intelligence is misreading human emotion, The Atlantic, https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/

Djemili, S., Longhi, J., Marinica, C., Kotzinos, D., Sarfati, G.E. (2014). What does Twitter have to say about ideology?. In *NLP 4 CMC: Natural Language Processing for Computer-Mediated Communication/Social Media-Pre-conference workshop at Konvens 2014*, Universitätsverlag Hildesheim.

Ekman, P. (1984). Expression and the nature of emotion. *Approaches to emotion*, pp. 319-344.

Essebbar, A., Kane, B., Guinaudeau, O., Chiesa, V., Quénel, I., Chau, S. (2021). Aspect Based Sentiment Analysis using French Pre-Trained Models. In *ICAART* (1), pp. 519-525.

Lara, E.A., Márquez, A.C., Fuentes-Rodríguez, C. (2016). Emotional argumentation in political discourse. *A gender-based approach to parliamentary discourse*, 68, pp. 129-159.

Longhi, J. (2013). Essai de caractérisation du tweet politique. *L'information grammaticale*, 136, pp. 25-32.

Longhi, J. (2020). Proposals for a discourse analysis practice integrated into digital humanities: theoretical issues, practical applications, and methodological consequences. *Languages*, 5(1), 5.

Ma, J., Li, X., Ren, Y., Yang, R., Zhao, Q. (2021). Landmark-Based Facial Feature Construction and Action Unit Intensity Prediction. *Mathematical Problems in Engineering*, https://www.hindawi.com/journals/mpe/2021/6623239

Mostowlansky, T., Rota, A. (2020). Emic and etic. *Cambridge Encyclopedia of Anthropology*, https://www.anthroencyclopedia.com/entry/emic-and-etic

Pike, K.L. (1954). *Language in relation to a unified theory of the structure of human behavior, part 1*. Glendale, Calif.: Summer Institute of Linguistics.

Quiros-Ramirez, M.A., Onisawa, T. (2015). Considering cross-cultural context in the automatic recognition of emotions. *International journal of machine learning and cybernetics*, 6(1), pp. 119-127.

Scherer, K.R., Clark-Polner, E., Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, 46(6), pp. 401-435. https://doi.org/10.1080/00207594.2011.626049

Simon, J., Bénédicte, T. (2016). Le poids argumentatif des tweets multimodaux dans les récits médiatique. In *Colloque international Médias numériques et communication électronique*, Klog, pp. 227-238.

Soysal, Ö. M., Shirzad, S., Sekeroglu, K. (2017). Facial action unit recognition using data mining integrated deep learning. In *2017 International Conference on Computational Science and Computational Intelligence* (CSCI), pp. 437-443.

Tenney, I., Das, D., Pavlick, E. (2019). BERT rediscovers the classical NLP pipeline. arXiv preprint arXiv:1905.05950.

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PloS one*, 12(5).

Models:

https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment?text=Madame+Le+Pen%2C+la+France+mérite+mieux+que+vous

https://huggingface.co/BaptisteDoyen/camembert-base-xlni?candidateLabels=colère%2C+dégout%2C+surprise%2C+tristesse&multiClass=false&text=Madame+le+Pen%2C+la+France+mérite+mieux+que+vous

Live Text Coverage of Political Events: Combining Content and Corpus-based Discourse Analysis

Hendrik Michael, Valentin Werner
University of Bamberg, Germany
E-mail: hendrik.michael@uni-bamberg.de, valentin.werner@uni-bamberg.de

Abstract

Live text (LT) has emerged as a web-native CMC alternative to traditional forms of live broadcasting. Through a combined content and corpus-based discourse analysis of the LT coverage of a major political event (the 2020 US presidential debates), the present study tests (i) how current LTs emphasize transparency and accountability, and (ii) how they are a form of journalistic communication that normalizes professional norms of objectivity in hybrid media settings. Political LT emerges as multi-layered and multi-authored discourse that places strong emphasis on accountability and disclosure transparency by updating and linking information, while maintaining the journalistic gatekeeping/gatewatching function. Linguistically, it is characterized by an informal tone but also by a continuation of traditional news media practices as regards objectivity, as instantiated by the salience of debate topics and political terms and – unlike the more widely studied sports LT – by a clear delineation of information from opinion and contextualization.

Keywords: online news, live text, live blogging, hybridity, political journalism, digital journalism, media studies

1. Introduction: (Digital) Journalism and Politics

Journalism is the main source of political information in deliberative democracies (McNair, 2009; Schudson, 2008) and it functions as gatekeeper and agenda-setter (Shoemaker et al., 2009), specifically during election periods (Schulz, 2011). In this context, television became the main channel through which the public receives political information (e.g. Farnsworth & Lichter, 2010; Gottfried et al., 2016). Previous analysis has shown that broadcast journalism is particularly important for voters as it puts campaigns into the public spotlight. A prime example for this are the presidential debates, a cornerstone of every campaign season (Kraus, 2020). However, the advent of hybrid media systems (Chadwick, 2017) during the last decade(s) has markedly altered how journalists cover politics and how audiences consume relevant content. An increasing number of recipients, and particularly younger segments of the audience, use online news and social media as their primary sources of political information now (Mitchell et al., 2020).

Political journalism today has defining characteristics such as multimediality, connectivity and hypertextuality (Nuernbergk & Neuberger, 2019), which foster media convergence and hybrid journalistic practices. Multimediality and connectivity allow cross-media production by which news content is effectively played out on different platforms and establishes connections with different target audiences. Connectivity means that online journalism follows and implements what other media are publishing, making news media producers gatewatchers that scan the flow of information in digital publics and provide context by linking and fact-checking (Bruns, 2018). By linking and updating (inaccurate) information online news media can create transparency and accountability in ways that TV and print media lack. Despite the evolution of these hybrid media practices, online journalism is also subject to a professional normalization. Studies by Singer (2005) and Lasorsa et al. (2012) have shown that long-established roles and practices of professional news media are still in operation. For instance, they shape to a significant degree how journalists share information by relying on media of record or their host media when providing links. In addition, journalists have been found to mostly adhere to standards of objective reporting by using impersonal and informative language.

A paradigmatic example of hybrid media practices, illustrating how political information is mediated and consumed, is live text (LT), alternatively labeled "live blogging". LT has emerged as a popular web-native alternative to TV and radio broadcasting for live reporting on pre-scheduled events with a limited duration (Thurman & Walters, 2013). This form of CMC is characterized by modularity (various content and navigational zones), multimodality (embedded multimedia content) and its dynamic nature (see Figure 1 for an example).

Usually, LT is multi-authored and reports on events in reverse chronological order. As the discourse emerges as the events reported upon unfold, it has also been termed a "text-in-motion" (Hauser, 2008) "text-in-process" (Chovanec, 2018) or "open news discourse" (Thorsen & Jackson, 2018), delineating it from traditional news practices.

To explore whether and in which forms the above-mentioned trends in hybrid media systems are empirically traceable in current journalism, the present study focuses on LT in political reporting (cf. Tereszkiewicz, 2014; Van Driel, 2020) and aims to illustrate current journalism practice in a CMC format.

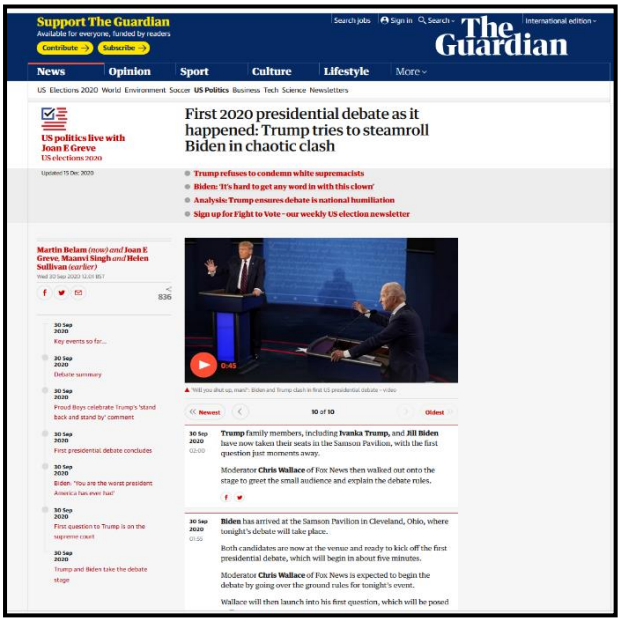


Figure 1: LT from *The Guardian*[1]

2. Data, Method, and Research Hypotheses

This study focuses on a specific form of CMC, namely political LT produced by professional journalists. Thus, the analysis relies on a purpose-built corpus of LT coverage of two televised US presidential debates (Donald Trump vs. Joe Biden), which both lasted for approximately 90 minutes and were screened on 29 September (73 million viewers) and 22 October 2020 (63 million viewers), respectively. Data were collected from four popular media outlets (*The Guardian*, *Daily Mirror*, *New York Times*, *Wall Street Journal*) to represent both British and US-American perspectives, to include news organizations associated with the tabloid and the quality press, and to represent media that have liberal and conservative editorial stances. The overall corpus size amounts to 61,490 tokens (excluding audience contributions through Tweets, etc.).

By comparison, the LT corpus compiled is small, but was designed for a close analysis of current media practice, taking into account the multimodal nature of LT as an understudied CMC manifestation. The LT contained in the corpus was created either by a single reporter (*Daily Mirror*) or by teams of up to 16 different journalists (*New York Times*). The difference in contributors is also reflected in the number of individual posts per reporting (ranging from $n = 35$ to $n = 319$), while the length of individual posts varied between 33 and 199 words. Generally speaking, a higher post frequency and a higher number of commentators was observed to correlate with shorter individual posts. The aforementioned findings are clearly indicative of varying practices in different media outlets.

To facilitate a discourse-oriented mixed-methods approach (O'Keeffe, 2006; Bednarek & Carr, 2020), data was stored both in its original HTML format to allow the consideration of multimodal aspects and manual annotation in *MAXQDA* (MAXQDA, 2019), as well as in TXT format to allow quantitative analysis (wordlists, *n*-grams) in the corpus software *AntConc* (Anthony, 2020). Manual data annotation was conducted by a research assistant and double-checked by the co-authors.

With the help of the aforementioned data and given the general characteristics provided in Section 1, the present study tackles the following broader hypotheses:

- H_1: Presidential debate LTs emphasize transparency and accountability by updating and linking information, while maintaining the journalistic gatekeeping/gatewatching function.
- H_2: Presidential debate LTs are a form of journalistic communication that normalizes professional norms of objectivity in hybrid media settings.

For the operationalization of H_1, we conduct a content analysis and, adapting categories established in media studies (e.g. Singer, 2005; Bruns, 2018), we annotate (i) sources of information used, such as political agents, media of record, ordinary people, etc., as well as (ii) indications of transparency and accountability in terms of updates, links, and quotes. For H_2, we test how criteria for journalistic objectivity, such as the lack of personal stance, fair representation, fair skepticism, identification of better arguments, and sticking to hard facts, etc., originally postulated in a classic study by Donsbach & Klett (1993), are reflected linguistically. This may materialize, for instance, in terms of salient content words, (non-)usage of personal pronouns, mental verbs, and reporting verbs.

3. Results

3.1. Content Analysis

The data from the content analysis suggest a clear pattern, namely that most information was sourced from media of record, host media, and political agents (Figure 2). Note that numbers for the last category would be much higher (>10) if statements by the presidential candidates, which naturally are very salient in the discourse (see also Section 3.2.1), were included as sources of information as well. The difference in sources from media of record between the first and the second debate can be explained by the fact that shortly before the first debate, the *New York Times* published an investigative exposé on Trump's tax record that was referred to repeatedly. Notably, only few scientific or public service organizations were referenced as sources of information, and neither are ordinary people often given a voice.

[1]https://www.theguardian.com/us-news/live/2020/sep/29/presidential-debate-latest-news-tonight-watch-trump-biden-taxes-coronavirus-updates

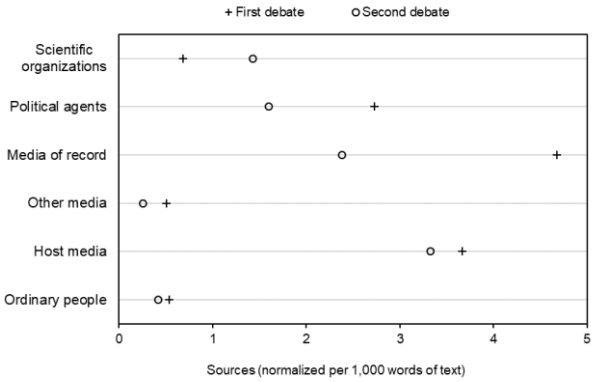


Figure 2: Source of information in political LT

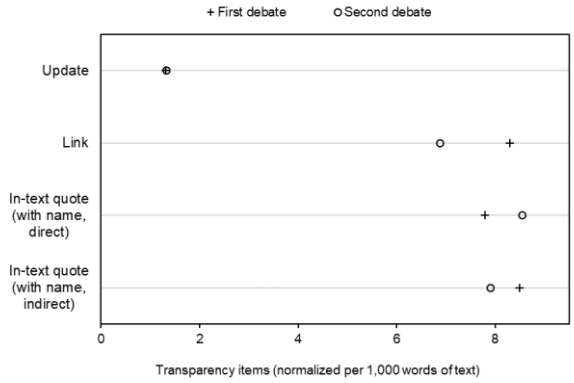


Figure 3: Transparency items in political LT

With respect to linking and updating, the data yield a mixed picture (Figure 3). Updates are relatively scarce, given that for example *The Wall Street Journal* and the *Daily Mirror* do not indicate updates at all, while this practice of creating transparency is only institutionalized by *The Guardian*. By contrast, linking emerges as a routine that is broadly used as an efficient strategy to create transparency and accountability. Similar rates of usage are observable for using named direct and indirect quotes to foster accountability. Unnamed quotes (not shown in Figure 3) are rare and are exclusively used in the *New York Times* LT.

3.2. Corpus Analysis

3.2.1. Content Words

Among the highly frequent (top 100) content words, the following categories emerge:

1. Names of participants: *Trump, Biden, Joe, Wallace, moderator*
2. Political discourse: *(First/presidential) debate, president, election, people, campaign, Americans, voters, candidates*
3. Debate topics: *Coronavirus, pandemic, (Supreme) court, health, states, US, China, climate, the Affordable Care Act*
4. Liveness: *Now, here, before, tonight*
5. Accountability: *Fact (check)*
6. Proclamatory character: *Say*
7. Personal stance: *(I) think*

Categories 1–3 demonstrate the fact-based/event-based nature of the LT coverage, while 4 is illustrative of the dynamic nature of the medium (see Section 1). 5 directly relates to the issue of accountability (see Section 2), while 6 and 7 are the most frequent items to report facts and personal stance, respectively, which merit a closer look (see further Section 3.2.2). In the corpus, *I think* is typically used to delineate opinions and value judgements by the reporters from information, as in *Adam, I think you've got it right here* (*New York Times*), simultaneously supporting the construction of an expert identity as a typical journalistic practice (Ekström & Lundell, 2011). In addition, *I think* serves the purpose to represent the vox pop, a strategy conducive to the objectivity principle of fair representation.

3.2.2. *N*-grams

Among the most frequent 3/4-grams, we find *Mr Trump/Mr Biden said*, often preceded (and/or followed) by a direct quote as in *Mr Trump said he wants "crystal clean water and air"*, which could be interpreted as the mere reporting of hard facts through a proclamatory verb.

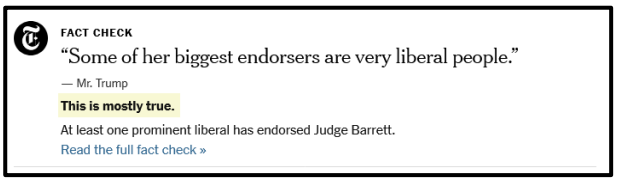


Figure 4: "Fact check" post with link from the *New York Times*[2]

In addition, *Mr Trump this is* (preceded by quote, followed by evaluative adjective) is highly frequent, as in *"I'm going to release them [his tax return forms] as soon as we can. I wan't to do it" - Mr Trump. This is misleading….* This recurrent pattern is indicative of the fact checking/contextualization function that fosters objectivity (see Figure 4 for another example as appearing in the LT).

3.2.3. Reporting Verbs

A dedicated look at reporting verbs (following the list from Garretson & Ädel, 2008) reveals that the most common neutral reporting verb lemma *say* is used with reference to either candidate in the same manner (see Table 1).
However, among other patterns, it also emerges (i) that *accuse*, an item usually employed by challengers rather than incumbents, surprisingly is used in relation to Trump, while *criticize* is associated with Biden and (ii) that the disfavoring item *claim* mostly collocates with statements by Trump.

[2]https://www.nytimes.com/interactive/2020/09/29/us/politics/debate-live-stream-analysis.html

LEMMA	Trump	Biden	LEMMA	Trump	Biden
SAY	164	166	CLAIM	59	2
TELL	16	17	PROMISE	5	1
ASK	5	3	ADD	11	14
CALL	8	17	CITE	4	1
TALK	20	12	ANSWER	3	3
ARGUE	9	8	RESPOND	12	25
ACCUSE	15	7	MENTION	4	2
SUGGEST	10	2	DEFEND	10	3
CRITICIZE	5	12	INSIST	8	3
SPEAK	4	7	ACKNOWLEDGE	5	1
NOTE	2	7	REPEAT	10	2

Table 1: Reporting verbs (lemmas)

4. Discussion and Conclusion

Overall, LT can be viewed as a clear instantiation of CMC that exemplifies media convergence (or hybridity), in which formerly separate communicative practices are consciously combined to make their points (Bateman et al., 2017). Its multi-layered and multi-authored discourse is illustrative of dynamic news creation through discourse (Karlsson, 2011), which is facilitated in a CMC environment. Arguably, this process orientation, which involves constant updating, creates accountability. Further support for H_1 was evident. Through gatewatching, LT relies on a mix of sources, including opinion statements and evaluation (Bruns, 2018), and thus creates "disclosure transparency" (Pantic et al., 2017). At the same time, it emerges that elite sources dominate and that journalists apparently continue to fulfill their gatekeeping or "filtering" function, which reduces the participatory transparency, very much despite the affordances of LT as an interactive medium.

While previous studies (Tereskiewicz, 2014) have diagnosed an overall informal tone in political LT, as regards objectivity and its linguistic representation, political LT appears to continue traditional news media practices concerning objectivity principles (cf. Donsbach & Klett, 1993), as stated in H_2. This is traceable linguistically, for instance pertaining to the salience of debate topics and political terms, as well as in terms of the delineation of information (*X said…*) vs. opinion (*I think…*) vs. contextualization (*This is exaggerated…*). However, also some bias in reporting verbs, occasionally boosted by evaluative items (*weirdly claims…*), could be observed in the data.

The latter finding notwithstanding, the present data are supportive of a general trend toward normalization of online news discourse (Singer, 2005; Lasorsa et al., 2012) and imply that matter-of-fact-style political LT should be clearly differentiated from other LT types, such as sports LT, where the merging of reporting, commenting and glossing and a dedicated infotainment function is much more explicit (cf. Werner, 2016, 2019).

Overall, by way of a case study, the present analysis was intended (i) to provide an insight into current mediatized journalistic language (Van Hout & Burger, 2017), and (ii) to highlight the potential of a combined media-linguistic approach toward political reporting through a from of CMC. Therefore, it may also inform the wider discussion in neighboring fields, such as communication and (digital) journalism studies.

5. References

Anthony, L. (2020). *AntConc* (3.5.9). Waseda: Waseda University.

Bateman, J., Wildfeuer, J. and Hiippala, T. (2017). *Multimodality: Foundations, research and analysis.* Berlin: Mouton de Gruyter.

Bednarek, M., Carr, G. (2020). Computer-assisted digital text analysis for journalism and communications research: Introducing corpus linguistic techniques that do not require programming. *Media International Australia*. doi.org/10.1177/1329878X20947124

Bruns, A. (2018). *Gatewatching and news curation: Journalism, social media, and the public sphere*. Frankfurt: Lang.

Chadwick, A. (2017). *The hybrid media system: Politics and power*. Oxford: Oxford University Press.

Chovanec, J. (2018). Participating with media: Exploring online media activity. In C. Cotter & D. Perrin (Eds.), *The Routledge handbook of language and media.* London: Routledge, pp. 505-522.

Donsbach, W., Klett, B. (1993). Subjective objectivity: How journalists in four countries define a key term of their profession. *Gazette*, 51, pp. 53-83.

Ekström, M., Lundell, Å.K. (2011). The joint construction of a journalistic expert identity in studio interactions between journalists on TV news. *Text & Talk*, 31(6), pp. 661-681.

Farnsworth, S.J., Lichter, R.S. (2010). *The nightly news nightmare: Media coverage of U.S. presidential elections, 1988–2008*. Lanham: Rowman & Littlefield.

Garretson, G., Ädel, A. (2008). Who's speaking? Evidentiality in US newspapers during the 2004 presidential campaign. In A. Ädel & R. Reppen (Eds.), *Corpora and discourse: The challenges of different settings.* Amsterdam: Benjamins, pp. 157-187.

Gottfried, J., Barthel, M., Shearer, E. and Mitchell, A. (2016). The 2016 presidential campaign - a news event that's hard to miss. *Pew Research Center – Journalism & Media*. www.journalism.org/2016/02/04/the-2016-presidential-campaign-a-news-event-thats-hard-to-miss/

Hauser, S. (2008). Live-Ticker: Ein Neues Medienangebot zwischen medienspezifischen Innovationen und stilistischem Trägheitsprinzip [Live ticker: A new media type between media-specific innovations and stylistic lag]. *kommunikation@gesellschaft,* 9(1), pp. 1-10.

Karlsson, M. (2011). The immediacy of online news, the visibility of journalistic processes and a restructuring of journalistic authority. *Journalism,* 12(3), pp. 279-295.

Kraus, S. (2020). *Televised presidential debates and public policy*. London: Routledge.

Lasorsa, D.L., Lewis, S.C. and Holton, A.E. (2012). Normalizing Twitter: Journalism practice in an emerging communication space. *Journalism Studies*, 13(1), pp. 19-36.

MAXQDA. (2019). *MAXQDA*. Berlin: VERBI Software.

McNair, B. (2009). Journalism and democracy. In K. Wahl-Jorgenson & T. Hanitzsch (Eds.), *The handbook of journalism studies*. London: Routledge, pp. 237-249.

Mitchell, A., Jurkowitz, M., Baxter Oliphant, J. and Shearer, E. (2020). Americans who mainly get their news on social media are less engaged, less knowledgeable. *Pew Research Center – Journalism & Media.* www.journalism.org/2020/07/30/americans-

who-mainly-get-their-news-on-social-media-are-less-engaged-less-knowledgeable/

Nuernbergk, C., Neuberger, C. (2018). *Journalismus online: Profession – Partizipation – Technisierung* [Journalism online: Profession – participation – technologization]. Heidelberg: Springer VS.

O'Keeffe, A. (2006). *Investigating media discourse*. London: Routledge.

Pantic, M., Whiteside, E. and Cvetkovic, I. (2017): Politics, conflict generate more live-blog comments. *Newspaper Research Journal,* 38(3), pp. 354-365.

Shoemaker, P., Vos, T. and Reese, S. (2009). Journalists as gatekeepers. In K. Wahl-Jorgenson & T. Hanitzsch (Eds.), *The handbook of journalism studies*. London: Routledge, pp. 73-87.

Schudson, M. (2008). *Why democracies need an unlovable press*. Malden: Wiley.

Schulz, W. (2011). *Politische Kommunikation: Theoretische Ansätze und Ergebnisse empirischer Forschung* [Political communication: Theoretical approaches and results of empirical studies]. Wiesbaden: Springer VS.

Singer, J.B. (2005). The political j-blogger: 'Normalizing' a new media form to fit old norms and practices. *Journalism,* 6(2), pp. 173-198.

Tereskiewicz, A. (2014). "I'm not sure what that means yet, but we'll soon find out": The discourse of newspaper live blogs. *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, 131, pp. 299-319.

Thorsen, E., Jackson, D. (2018). Seven characteristics defining online news formats: Towards a typology of online news and live blogs. *Digital Journalism*, 6(7), pp. 847-868.

Thurman, N., Walters, A. (2013). Live blogging: Digital journalism's pivotal platform. *Digital Journalism*, 1(1), pp. 82-101.

Van Driel, M. (2020). "It makes it more real": A comparative analysis of Twitter use in live blogs and quotations in older news media from a reader response perspective. In C. Tagg & M. Evans (Eds.), *Message and medium.* Berlin: Mouton de Gruyter, pp. 306-328.

Van Hout, T., Burger, P. (2017). Mediatization and language of journalism. In O. García, N. Flores & M. Spotti (Eds.), *The Oxford handbook of language and society.* Oxford: Oxford University Press. doi.org/10.1093/oxfordhb/9780190212896.013.9

Werner, V. (2016). Real-time online text commentaries: A cross-cultural perspective. In C. Schubert & C. Sanchez-Stockhammer (Eds.), *Variational text linguistics: Revisiting register in English.* Berlin: Mouton de Gruyter, pp. 271-305.

Werner, V. (2019). A multimodal analysis of football live text commentaries. In M. Callies & M. Levin (Eds.), *Corpus approaches to the language of sports: Texts, media, modalities.* London: Bloomsbury, pp. 201-240.

Hashtags in English Tweets About the New World Order

Natalia Mora-López
Complutense University of Madrid
E-mail: nataliamora@ucm.es

Abstract

The New World Order conspiracy theory can be found in the media in relation to current politics, social events, institutions and the health situation and measures taken. This study aims to shed some light on how this conspiracy theory is represented on Twitter by analysing and classifying the meanings of the 100 most common hashtags in 26,038 tweets retrieved during the summer of 2020 through the free add-on for Google Spreadsheets Tweet Archiver. The results show that the meanings of the hashtags can be divided into five categories: (1) Actions and Reactions, (2) Entities, (3) Health, (4) Politics, Society and Religion, and (5) Other.

Keywords: conspiracy theory, hashtags, New World Order, Twitter

1. Introduction

Social media have been associated with beliefs in conspiracy theories and seen to be a key player in their dissemination as well (Stempel et al., 2007; Jamieson & Albarracín, 2020; Enders et al., 2021). They also provide a wide, open and free source for the creation of corpora, which allows quantitative and qualitative linguistic and non-linguistic analyses of what is posted in them. This is the case of Twitter, whose comments typically include a convention, the hashtag, that indicates that the tweet is related to other tweets because they share the same topic. This conversational tagging uses a hash before the name of the label used to classify the comment. Hashtags appeared on Twitter through community use (Huang et al., 2010) and may be considered a folksonomy or form of social tagging (Vander Wal, 2007). They are mainly used in final position, taking the whole comment as scope (e.g., "All in one package #NWO"), or integrated as part of the comment itself by using the hash before words that are functioning as such in a sentence (e.g., "#Soros the #NWO Devil").

Hashtags have been studied from several perspectives. Sociologically, hashtags have been shown to organise discussions on, for example, politics (Heverin & Zach 2011) or to create publics (Bruns & Burgess 2011) as well as playing key roles in political activism and marketing and transforming digital culture (Bernard, 2019). Computationally, researchers have studied how hashtags spread (Chang 2010), which hashtags may be automatically recommended to users to include in their tweets (Mazzia & Juett, 2009; Jeon, Jun & Hwang, 2014; Tran, Hwang & Nguyen, 2018), whether they are positive or negative (Wang et al., 2011), or how they help to classify tweets (Antenucci et al., 2011). Linguistically, the functions and metafunctions of hashtags have been analysed (Dickinson, 2013; Zappavigna 2015), as well as their word-formation processes (Caleffi, 2015) and the concept and definition of hashtags themselves (La Rocca, 2020). The analysis of hashtags in the context of posts related to conspiracy theories may help understand what connections are made between certain concepts or social issues to gain better understanding of how conspiracy theories work.

2. Aims

This study aims to determine the most common meanings of the most frequent hashtags used in tweets related to the New World Order conspiracy to shed some light on what kind of hashtags may become powerful in this context. Specific categories into which the hashtags can be classified will be created for this purpose.

3. The New World Order conspiracy theory

The key idea in this conspiracy theory is that a secretive group of people are leading the world towards a totalitarian government (Steward, 2002). Conspiracists also believe that this group wants to reduce world population and that governments spy people through technology. The spread of Covid-19 and the reaction that has been produced world-wide has also been seen as part of the conspiracy, since conspiracists have found connections between the virus and the above-mentioned ideas. Examples of these connections are the idea that technology can be used to track infected people and then control what the population does, or that the origin of the virus is in China, a communist country ruled by a dictator.

4. Approach, corpus and methodology

This study uses a corpus-based approach to the classification of the most common hashtags found in the collected tweets. Hashtags will be analysed quantitatively and qualitative regarding their meaning. However, this meaning will be obtained in the context of the conspiracy theory in which they are used.

The hashtags analysed belong to a corpus of 26,038 English tweets posted from 20th June 2020 to 22nd September 2020, related to the conspiracy theory that involves the establishment of a New World Order. These tweets were automatically collected via Tweet Archiver, a free add-on for Google Spreadsheets, by retrieving tweets including the hashtags #NewWorldOrder and #NWO, using these hashtags as search terms in two different searches and removing duplicates that included both of them. The problems and issues found in the retrieval were discussed by Mora-López (2020).

Frequency lists of the hashtags were obtained via Sketch

Engine through two different searches, due to the limitations automatic retrieval finds when classifying strings starting with "#" as a word. Often, Sketch Engine retrieves "#" in isolation as an independent word, leaving the actual hashtag as the following independent word. For example, it retrieved #plandemic, #nwo, #who as single words, but "#" was a separate item from "5g" (#5g), "4life" (#4life) or "rt" (#rt). The first search consisted in using the Wordlist tool to find words and non-words matching the regular expression "#.+" to retrieve any lexical items starting with "#". The second one, which had the purpose of addressing the limitation, used the N-gram feature to search for 2-grams matching the regular expression "#\ .*" to retrieve any words and non-words which were preceded by "#". The two lists were combined to extract the most frequent hashtags, which were then analysed and classified manually. A total of 20,122 different hashtags were obtained with a total of 154,978 occurrences.

5. Results and discussion

The 100 most frequent hashtags were manually analysed by the author of this paper to find the common topics they dealt with. These topics were grouped into general categories that represented the most important issues or concepts linked to the conspiracy theory. This set is broad enough to comprise all tweets but at the same representative enough to shed light on the topics dealt with. Additionally, these broad categories were also subdivided into more specific categories inside for a more accurate description of the meaning relationships found. The categories were created in an inductive way, that is, they were obtained from the meanings of the hashtags retrieved and the contents of the tweets. The full account of hashtags can be seen in Appendix 1. It must be noted that hashtags were either used as part of the text or at the end of the tweet.

5.1. Actions and Reactions

These included hashtags related to either carrying something out and requesting to do so, or refusing to do it. The subcategories are *Call to action* (#wakeup, #protest), *Opposition* (#nomasks, #nonewnormal) and *Request* (#freedom, #truth).

5.2. Entities

This category included *Country* (#usa, #china), *Institution or Organisation* (#un, #savethechildren), *Person* (#trump, #billgates), *Qualifier*, this being entities modified inside the same hashtag (#plandemic, #covididiots) and *Social group* (#illuminati, #qanon). These hashtags were used in tweets that addressed any of these entities and said something about them.

5.3. Health

There were only two subcategories in Health: Illness (#covid19, #coronavirus) and Measures (#masks, #vaccine). The appearance of these hashtags is a result of the introduction of the COVID-19 pandemic as another element inside the New World Order conspiracy theory. This is due to the fact that it is related to several aspects that define the conspiracy, as mentioned in section 3.

5.4. Politics, Society and Religion

This category groups different but interconnected concepts: *Control* (#depopulation, #deception), *Economy* (#thegreatreset), *Event, Plot or Project* (#agenda21, #obamagate), *Ideology or Political orientation* (#globalists, #democrats), *Media* (#news, #breakingnews), *Religion* (#cabal, #luciferian), *Slogan* (#blacklivesmatter, #maga) and *Technology* (#5g).

5.5. Other

Finally, some hashtags were not found to belong to any other the previous categories and did not totally shed any light on meanings related to the conspiracy theory. These were classified as Other. Inside, there were three subcategories: *Book* was specifically created due to the high number of occurrences of hashtags referring to a single fiction book that uses the New World Order conspiracy (and additional fictional elements) as setting (#fiction, #kindle). By linking the conspiracy hashtags to a fictional book, these posts do not deal with any issues of real-world conspiracy. Actually, the high number of occurrences for this category was due to the repetition of the same tweet with commercial purposes. The subcategory *Marketing* was used for a single hashtag (#xyz), which users seem to believe will position their tweets better, but which is not related to the conspiracy either. Thirdly, the subcategory *MetaNWO* comprised the hashtags naming the NWO itself, that is, the two hashtags used for the searches, so at least one of them appeared in every tweet (#nwo, #newworldorder).

5.6. Discussion

A summary of the categories and the number of hashtags found is presented in Table 1. This table shows the number of occurrences for each category, that is, how many times the hashtags in that category were used, and the percentage that these occurrences represent in the total amount of hashtags retrieved. As can be seen, the 100 most common hashtags add up to 46.46% of all the hashtags used.

Cat.	No. Occurrences	% in total #tags
A/R	4,251	2.74%
E	13,770	8.89%
H	8,427	5.44%
P/S/R	13,998	9.03%
O	31,555	20.36%
Total	72,001	46.46%

Table 1: Number of occurrences of the categories.

Without taking into account the category Other, which was mainly based on the *Book* and the *MetaNWO* results, it must be noted that the most frequent categories are Politics, Society and Religion, and Entities.

With respect to Politics, Society and Religion, the most common subcategories were *Event, Plot or Project* (5,099 occurrences), *Media* (2,472 occurrences), *Ideology or Political orientation* (2,205 occurrences) and *Slogan* (2,068 occurrences). The first one mentioned political and social events, plots or projects that conspiracists consider somehow related to the conspiracy. For example, they talk about secret agendas or international events like Global Goal Unite that they think support and contribute to globalism. Conspiracists consider these events to be linked to or caused by the agents in the conspiracy. Hashtags classified as *Media* were basically used to call users attention on specific events. Hashtags in the *Ideology or Political Orientation* subcategory were more intrinsically related to the conspiracy and named types of governments, typically oppressive or considered oppressive by the conspiracist (thus, related to *Control*). Finally, the *Slogans* found were related to the Q-Anon group, the Black Lives Matter movement and Trump's campaign. The overlapping of fields made it difficult to separate these hashtags into separate categories. For example, the Black Lives Matter movement is at the same time a social and a political event, criticised by conspiracists, who indicate in the tweets that it is related to George Soros, democrats, socialism, globalism, or the secret agenda, among others.

Regarding Entities, it is based on naming agents that conspiracists relate to the New World Order. These entities are typically accused of playing a negative role in the conspiracy, including Bill Gates, George Soros, Jeffrey Epstein, the Rothschild family, the World Health Organisation and the United Nations, among others. The only mentioned agent who was playing a positive role, according to the conspiracists, was Donald Trump.

Because of the time when these tweets were retrieved (the summer of 2020), the number of Health results associated with the New World Order is also very high, due to the mass presence of Covid-19 in the media and how conspiracists included it as part of the conspiracy.

Finally, although Actions and Reactions is the least frequent group in the analysed hashtags, it is also a key element in the conspiracy. In this category, users most commonly address readers directly by urging them to do something, usually "waking up", that is, unmasking the lies they are being told and embracing the truth conspiracists are showing. Another frequent use of hashtags is to name a positive concept that they think they do not currently have (e.g., freedom, peace, justice, equality) or they would like to spread (e.g., love, truth). Less frequently, they also show their opposition to the new social situation government by health measures or to specific measures.

Among the five categories, Other stands for those hashtags which do not provide any further information about the conspiracy itself, including the fictional book using the conspiracy as setting, the use a of an unrelated hashtag for positioning purposes and the hashtags used for the search. However, the remaining four categories (Actions and Reactions; Entities; Health; and Politics, Society and Religion) do actually show trends in tweets and what the main topics and subtopics are considered to be related to the conspiracy.

6. Conclusions

This study has analysed the main topics into which the 100 most frequent hashtags in tweets about the New World Order conspiracy theory can be divided. It has shown that these topics are the following ones: Actions or Reactions; Entities; Health; Politics, Society or Religion; and Other. It has been seen that, apart from Other, which basically comprised the hashtags used for the retrieval (#NewWorldOrder and #NWO) and hashtags not directly related to the conspiracy as such but to a fiction book whose setting takes elements from it, the most common categories were Politics, Society or Religion, and Entities. These categories were used to mostly refer to agents and events that play a role in the conspiracy. The Health category has an important representation in the corpus because of the Covid-19 situation which overlapped with the retrieval of the tweets in the summer of 2020 and the introduction of this virus in the conspiracy. Finally, Actions and Reactions were used, in a less frequent way, by conspiracists to call readers to action and express their personal opposition to a situation.

It can be then concluded that these hashtags, and consequently the tweets they belong to, focused on those entities playing roles in the conspiracy, and are to a great extent influenced by the moment when they are posted, thus referring to specific contemporary socio-political events, including the global health situation. Additionally, they make also use of the interactive nature of social media to urge readers to abandon their wrong thoughts or their ignorance and convince them of their own.

7. References

Antenucci, D., Handy, G., Modi, A., and Tinkerhess, M. (2011). Classification of tweets via clustering of hashtags. *EECS*, *545*, 1-11.

Bernard, A. (2019). *Theory of Hashtag*. Cambridge: Polity Press.

Bruns, A. and Burgess, J. E. (2011). The use of Twitter hashtags in the formation of ad hoc publics. In *6th European Consortium for Political Research General Conference*, August 25-27, University of Iceland, Reykjavik.

Caleffi, P. M. (2015). The "Hashtag": A new word or a new rule?. *SKASE Journal of Theoretical Linguistics*, 12, pp. 46–69.

Chang, H. C. (2010). A new perspective on Twitter hashtag use: Diffusion of innovation theory. *Proceedings of the American Society for Information Science and Technology*, 47(1), pp. 1-4.

Dickinson, P. (2013). "B/w U & me": The functions of formulaic language in interactional fiscourse on Twitter. *The Linguistic Journal*, 7(1), pp.7-38.

Enders, A. M., Uscinski, J. E., Seelig, M. I., Klofstad, C. A., Wuchty, S., Funchion, J. R., Murthi, M. N., Premaratne, K., and Stoler, J. (2021). The relationship between Social media use and beliefs in conspiracy theories and misinformation. *Political Behavior*, pp. 1-

24.

Heverin, T. and Zach, L. (2010). Microblogging for crisis communication: Examination of Twitter use in response to a 2009 violent crisis in the Seattle-Tacoma, Washington area. In *Proceedings of the 7th International ISCRAM Conference*. Seattle, Washington.

Huang, J., Thornton, K. M. and Efthimiadis, E. N. (2010). Conversational tagging in twitter. In *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, pp. 173-178.

Jamieson, K. H., and Albarracín, D. (2020). The relation between media consumption and misinformation at the outset of the sars-cov-2 pandemic in the us. *The Harvard Kennedy School Misinformation Review*, 1(2), pp. 1–22.

Jeon, M., Jun, S., and Hwang, E. (2014). Hashtag recommendation based on user tweet and hashtag classification on twitter. *International Conference on Web-Age Information Management*. Springer, Cham, pp. 325-336.

La Rocca, G. (2020). La fuerza de un signo: Perspectivas teóricas para el análisis de los hashtags. *Barataria: revista castellano-manchega de ciencias sociales*, 27, pp. 46-61.

Mazzia, A., and Juett, J. (2009). Suggesting hashtags on twitter. *EECS 545m, Machine Learning, Computer Science and Engineering, University of Michigan.*

Mora-López, N. (2020). Issues and uses of the retrieval of tweets for linguistic analysis: a contrastive case study on a set of four Spanish and English hashtags**.** *Working with Corpora and Digital Tools in Language and Translation Studies: Practical Issues and Methodological Challenges. International Webinar WocDit 2020*. Faculty of Philology, Complutense University of Madrid, 19-28 October 2020.

Stempel, C., Hargrove, T., and Stempel, G. H. (2007). Media use, social structure, and belief in 9/11 conspiracy theories. *Journalism & Mass Communication Quarterly*, 84, pp. 353–372.

Stewart, C. J. (2002). The master conspiracy of the john birch society: From communism to the new world order. *Western Journal of Communication,* 66(4), pp. 423-447.

Tran, V. C., Hwang, D. and Nguyen, N. T. (2018). Hashtag recommendation approach based on content and user characteristics. *Cybernetics and Systems*, 49(5-6), pp. 368-383.

Vander Wal, T. (2007) Folksonomy Coinage and Definition, http://vanderwal.net/folksonomy.html, date accessed: 10 March 2021.

Wang, X., Wei, F., Liu, X., Zhou, M. and Zhang, M. (2011). Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. *Proceedings of the 20th ACM international conference on Information and knowledge management*, pp. 1031-1040.

Zappavigna, M. (2015). Searchable talk: The linguistic functions of hashtags. *Social Semiotics*, 25(3), pp. 274-291.

Appendix 1. Full list of hashtags per category

Actions / Reactions			**4251**
Call to action	*2044*	*Request*	*1801*
#wakeup	1460	#freedom	506
#wakeupamerica	408	#truth	342
#protest	176	#peace	265
Opposition	*406*	#love	263
#nomasks	232	#justice	216
#nonewnormal	174	#equality	209

Entities			**13770**
Country	*753*	*Qualifier*	*5687*
#usa	270	#plandemic	2244
#china	251	#covidiots	808
#america	232	#scamdemic	678
Institution or Org.	*1563*	#covid1984	403
#un	516	#covidhoax	337
#who	438	#newnormal	228
#id2020	255	#newworld	219
#oneworldonenation	182	#policestate	213
#savethechildren	172	#falseflag	191
Person	*3268*	#fakenews	187
#trump	1369	#dystopia	179
#billgates	765	*Social group*	*2499*
#trump2020	410	#illuminati	991
#soros	355	#qanon	880
#rothschild	199	#qanons	258
#epstein	170	#q	187
		#anonymous	183

Health			**8427**
Illness	*7067*	*Measures*	*1360*
#covid19	3230	#masks	382
#covid	1946	#vaccine	367
#coronavirus	1891	#vaccines	353
		#lockdown	258

Other			**31555**
Book	*3729*	*Marketing*	*317*
#fiction	698	#xyz	317
#bookboost	667	*MetaNWO*	*27509*
#iartg	632	#nwo	21440
#endtimes	533	#newworldorder	6069
#prophecy	274		
#kindle	210		
#suspense	205		
#endofdays	176		
#spiritualwarfare	174		
#mustread	160		

Politics, Society and Religion			**13998**
Control	*628*	*Media*	*2472*
#depopulation	260	#news	822
#deception	186	#breaking	725

#psyop	182	#breakingnews	700
Economy	*208*	#msm	225
#thegreatreset	208	*Religion*	*1041*
Ev. Pl. & Pr.	*5099*	#thegreatawakening	314
#agenda21	1429	#cabal	229
#agenda2030	796	#markofthebeast	169
#deepstate	733	#greatawakening	166
#pizzagate	417	#luciferian	163
#conspiracy	289	*Slogan*	*2068*
#obamagate	263	#wwg1wga	701
#adrenochrome	246	#blm	533
#globalgoalunite	226	#maga	479
#lrd	185	#blacklivesmatter	355
#mkultra	177	*Technology*	*277*
#event201	173	#5g	277
#bmsproject	165		
Ideology or P.O.	*2205*		
#democracy	437		
#globalists	315		
#tyranny	261		
#antifa	207		
#communism	204		
#nazi	203		
#globalist	202		
#globalism	198		
#democrats	178		

Emotion Recognition and Sarcasm Mining Using Rule-Based and Deep Neural Networks

Abdelouafi El Otmani, Trang Lam, Julien Longhi
CY Cergy Paris Université - AGORA
E-mail: elotmani@gmail.com, doantrang1110@gmail.com, julien.longhi@cyu.fr

Abstract

Opinion and emotion recognition is a well-established task, with many standard data sets and well-developed methodologies. Nevertheless, French research suffers from a lack of annotated, publicly accessible resources to conduct experiments, despite a very large French-speaking population[1]. To address this lack, we collected a French tweet corpus from March 2020 based on the trending hashtags #confinement, #chloroquine, #vaccination, #stopcovid. The focus of this research is two-fold. The first objective is to analyze the public's reaction to Covid-19 topics such as confinement, medical care, and vaccination. The second objective is to propose a mixed approach of rule-based and deep learning to emotion recognition [1] and sarcasm mining [2]. More concretely, our procedure consists of two steps: 1) Due to the large of volume data and the lack of quality labeled datasets, we manually defined a set of rules to pre-annotate tweets, including extract six basic emotions [3] (anger, disgust, fear, happiness, sadness, and surprise) , identify the sentiment polarity and detect sarcasm. 2) Next, the pre-annotations are verified by a human and are used to train models. Different models are developed for this study such as CNN [4], LSTM + Attention [5] and Bert model [6]. The usefulness of these models can be explained by the fact that their representations robustly encode lexical and syntactic information. However, much work remains to be done to understand the extent to which we could adapt existing models to our complex problems. Thus, diverse techniques such as deconvolution, attention mechanism, and integrated gradient [7] are also used to evaluate the outcomes of these models.

Keywords: sentiment analysis, emotion recognition, sarcasm mining, rulebased, deep learning, covid19

References

Pecore, S. (2019). Analyse des sentiments et des émotions de commentaires complexes en langue française. Linguistique. Université de Bretagne Sud, Français. ffNNT:2019LORIS522ff. fftel-02903247.

Hernández Farias, D.I., Rosso, P. (2017). Chapter 7 - Irony, Sarcasm, and Sentiment Analysis, Editor(s): Federico Alberto Pozzi, Elisabetta Fersini, Enza Messina, Bing Liu, *Sentiment Analysis in Social Networks*, Morgan Kaufmann, pp. 113-128.

EKMAN, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6:3-4, 169–200. DOI : 10.1080/02699939208411068.

Kim, Y. (2014). *Convolutional Neural Networks for Sentence Classification*. EMNLP.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … & Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Sundararajan, M., Taly, A., & Yan, Q. (2017, July). Axiomatic attribution for deep networks. In *International Conference on Machine Learning* (pp. 3319-3328). PMLR.

Using Keystroke Logs to Analyze CMC

Erin Pacquetet
University at Buffalo
E-mail: erinmorr@buffalo.edu

Abstract

Computer-mediated communication is primarily produced through typing on a keyboard. whether it is a physical or a virtual one. Research studies often focus on the output of typing rather than the steps involved in production. Keystroke logging, the technique used to record typing production, redefines writing as a non-linear and spatiotemporal activity. Written data is static and shows the finished text. Keystroke data is dynamic. On top of all the static features that can be analyzed in a text, typing gives insights into how this text is produced, through dynamic metrics such as typing speed, pauses, and edits. Therefore, the study of typing helps uncover the processes to better understand the relationship between what is being typed and how it is typed.

In this study, I first go over the techniques used to collect and process keystroke logs for linguistic analysis, explaining how typing data is collected and how it should be formatted for better results, and how one can use Part of Speech tagging and Language models to parse keystroke data. Then, I present some of the key dynamic features that are specific to typing. Among those, we can cite pause location and length, which can highlight areas of the text that require more intense planning and are more difficult to produce, as well as make visible which sequences of words are processed and planned at once. Backspaces make production errors visible through revision and repair processes, which helps to understand the thought processes that occur during text production. Finally, typing speed, a correlate of fluency, enables us to evaluate how much of an obstacle to production the typing medium is. This research is a first step towards developing a comprehensive analysis method to use keystroke logs as a new layer for investigating dynamic text production in computer-mediated communication.

Keywords: typing, production, methodological

References

Chukharev-Hudilainen, E. (2014). Pauses in spontaneous written communication: A keystroke logging study. *Journal of Writing Research*, 6(1), 61.

Cislaru, G., & Olive, T. (2018). *Le processus de textualisation: analyse des unités linguistiques de performance écrite*. De Boeck Supérieur.

Leijten, M., Van Horenbeeck, E., & Van Waes, L. (2019). Analysing keystroke logging data from a linguistic perspective. In *Observing writing* (pp. 71-95). Brill.

Plank, B. (2016). What to do about non-standard (or non-canonical) language in NLP. arXiv preprint arXiv:1608.07836.

Working with Socially Unacceptable Discourse Online: Researchers' Perspective on Distressing Data

Kristina Pahor de Maiti*, Darja Fišer*♦‡

*Faculty of Arts, University of Ljubljana, Slovenia
♦Jožef Stefan Institute, Ljubljana, Slovenia
‡Institute of Contemporary History, Ljubljana, Slovenia
E-mail: kristina.pahordemaiti@ff.uni-lj.si, darja.fiser@ff.uni-lj.si

Abstract

This paper reports on the findings on annotators' and researchers' emotional and mental reactions triggered by work performed on distressing data, such as socially unacceptable discourse (SUD), discusses risk factors and possible mitigation measures. The data was obtained through a detailed self-observation study of individuals involved in the annotation and analysis of a corpus of socially unacceptable Facebook comments. The data shows that SUD comments are perceived as distressing which can negatively impact annotators' and researchers' well-being and their work. In addition, there are a number of risk factors that can further aggravate the situation, such as the length of immersion in the data and particular CMC characteristics. To counteract these effects, several mitigation measures are proposed, such as preparatory training and (in)formal debriefing. The findings clearly show that the domain of linguistics is not exempt from research hazard and that the well-being of the entire research team should be considered in the risk-assessment.

Keywords: hate speech, socially unacceptable discourse, distressing data, manual annotation, self-observation study, emotional labour, computer-mediated communication, Facebook

1. Introduction

Social media are becoming increasingly exploited as a rich and varied data source which can include distressing data. One such type of data is socially unacceptable discourse (SUD) which encompasses various offensive and aggressive communication practices including direct and indirect threats, offenses, incitements to violence, as well as negative generalizations, stereotype judgements, obscenities and provocations which negatively impact the communication (Vehovar et al., 2020). SUD is also known to negatively impact individuals and society as a whole (Asquith, 2010; Delgado, 2019). Less attention, however, has been paid to how it affects the research team. The aim of this paper is to investigate to what extent the research team perceives SUD comments as disturbing and to discuss possible mitigation measures.

2. Background

Data, including accounts of sensitive or traumatizing experience, can be emotionally burdening for the research team members (Dickson-Swift et al., 2008; Hanna, 2019), and can, in certain cases, even lead to developing symptoms similar to posttraumatic stress disorder (Figley, 1995; Moran & Asquith, 2020). The growing body of work increasingly highlights the need for research-related risk assessment, not only for research participants but also for members of the research team (cf. McCosker et al., 2001). The reason for this is that working with distressing data, such as police hate crime and suicide reports or (online) narratives of infertility or abuse (cf. Fincham et al., 2008; Hanna, 2019; Jackson et al., 2013; Moran & Asquith, 2020), may require intensive emotional labour. This is further supported by studies which invalidate the common belief that online data – often perceived as a detached form of communication – does not trigger an interpersonal connection and emotional response, which is a prerequisite for experiencing distress. However, research shows that online data, in fact, also necessitates active reading and absorption that can quickly lead to emotional saturation (Grant, 2018; Hanna, 2019). The hazard for the researchers' well-being can be importantly reduced by considering the possible risk factors. Among the most important ones, Moran and Asquith (2020) highlight "the amount, duration and type of exposure, as well as a person's life history and current circumstances". This outlines the need for comprehensive mitigation measures which should cover the entire research cycle and include adequate preparation, training, supervision as well as formal and informal debriefing administered in a regular and systematic fashion (Dickson-Swift et al., 2008; Fahie, 2014; Moran & Asquith, 2020).

3. Study Design

3.1. Research questions and hypotheses

Based on previous research and personal experience linked to possibly upsetting impact of SUD data on researchers and annotators (hereafter jointly referred to as *annotators*), we have formed the following research problems (RP) and hypotheses (H).

- RP1: REACTION TO SUD
  - H1.1: Annotators perceive working with socially unacceptable online comments as emotionally burdening (cf. Hanna, 2019; Stern, 2003).
  - H1.2: The distressing nature of SUD comments can impair annotators' objectivity in the work process (cf. Hanna, 2019).
  - H1.3: CMC linguistic features influence annotators' perception of the level of social unacceptability of the comments.
  - H1.4: Longer immersion in the data is linked to a more pronounced upset in annotators (cf. Moran & Asquith, 2020).

- H1.5: Data is more triggering for annotators with a personal connection to the topic in focus (cf. Moran & Asquith, 2020).

- RP2: MITIGATION MEASURES
  - H2.1: Initial training and annotation guidelines do not suffice in ensuring annotators' well-being in the work process.
  - H2.2: Most annotators seek additional support from colleagues, friends and family members (cf. Dickson-Swift et al., 2008).

3.2. Research instrument

The data for this study was obtained through a detailed online survey which was conducted in Slovene[1]. In addition to questions collecting sociodemographic information, the survey included 26 open- and close-ended questions and sub-questions, which were developed based on the related work and covered the perceived impact of working with SUD comments on annotators' well-being and performance. The survey also included 17 close-ended questions adapted from the *Secondary Traumatic Stress Scale* (STSS) (Bride et al., 2004), which is a tool to measure the symptoms associated with indirect exposure to a traumatic event. The results of this last part are not discussed in this paper due to length restrictions.

The online survey was sent to all 35 annotators and linguistic researchers who worked on the FRENK corpus. It consists of Facebook comments on the topic of migrants and LGBTQ+ in Croatian, English, French and Slovene which were manually annotated for the type and target of SUD, e.g. type: threat, offense, etc.; target: lgbt, migrant, commentator, etc. (Ljubešić et al., 2019). The corpus contains approximately 10.000 comments (roughly half per topic) per language.[2,3]

The response rate was 71% (25 unique and complete surveys were returned by the specified deadline), which is considered an adequate response rate (Baruch & Holtom 2008). Participation was voluntary, but the respondents were compensated for their time with a symbolic gift (cinema ticket). Consent was obtained from the respondents and answers were anonymized.

3.3. Sample

Sample characteristics. Out of 25 respondents, all worked as annotators, while three also participated as researchers. Table 1 shows that the majority were women aged 20–30, almost equally divided between those with BA or MA degree. In contrast, men, who represented one third of the respondents and were aged 20–35, predominantly held a BA degree. The respondents' field of study was in the Humanities (languages, translation studies, ethnology, sociology, history) or Social Sciences (journalism studies, communication studies), and none had previous experience with annotation/research task including distressing data.

	%	AGE (%)			EDU (%)	
		20-25	25-30	30-35	BA	MA
F	76	32	68	0	58	42
M	24	50	33	17	83	17

Table 1: Sample demographic information with education level.

As shown in Table 2, two thirds of the annotators worked with the data in short to medium time spans (1–5h in a row), one sixth annotated in long sessions (6 or more hours per day), while one fifth used different approaches in different campaigns (usually starting with shorter annotating sessions and prolonging them towards the deadline).

	ANNOT. PACE (h/d)				CAMPAIGN (#)		
	1–2	3–5	6+	Mix	1	2	3
%/RESP	32	28	16	24	36	28	36

Table 2: Percentage of respondents with regard to hours spent annotating per day and the number of annotation campaigns in which they participated.

4. Results and Discussion

This section reports and discusses the quantitative and qualitative results of the introspective survey obtained with basic descriptive statistical methods and thematic analysis.

4.1. Reaction to SUD

Emotional response to SUD. The data confirms *Hypothesis 1.1* stating that SUD online is perceived as distressing by the annotators, since almost 90% of the respondents reported some kind of predominantly negative emotional reaction to the comments. Among the most common emotions listed, we found *anger* (56% respondents), *sadness* (44%) and *bewilderment* (40%). These were followed by *disgust/resentment* (28%), *disappointment* (25%), *frustration* and *helplessness* (20% each), and *boredom* and *indignation* (16% each). The data shows that most respondents (73%) experienced a short- or medium-term emotional reaction (i.e., which stopped with the end of the campaign), while only some (≈27%) reported that their emotional reaction continued even after the end of the campaign. The importance of emotional risk assessment for the research group is supported by the fact that, although the majority experienced mild reactions (64%), there is a considerable share of those that reported having strong emotional responses to the comments (36%).

Mental response to SUD. Almost 65% of the respondents stated that they felt an obligation to supress their emotions, mostly in order to ensure objectivity of their annotations, but also to ward off the unpleasant negative feelings. Furthermore, respondents reported to have experienced

[1] The survey can be obtained from the first author upon request.

[2] This count includes the first author of this paper since she participated in the project as an annotator and researcher.

[3] The FRENK project: https://nl.ijs.si/frenk/english/

various negative mental reactions, such as quickly becoming warry of the work, struggling to stay concentrated or adopting a cynical attitude towards the comments. Based on the results, we thus confirm *Hypothesis 1.2* stating that SUD comments may impair objectivity of annotators and researchers. However, this is only true if reactions to such data is not consciously addressed (cf. *Hypothesis 1.5* in the next section). Contrary to the common research practice, when working with distressing data, disregarding emotions might not be in the best interest of the study (cf. Dickson-Swift et al., 2009). In fact, Moran and Asquith (2020) advocate for researchers' reflexivity which includes noting and understanding one's emotions and personal vulnerabilities during research, since, when not addressed, this can lead to blind spots that undermine a researcher's capacity for analysis. Similarly, our data shows that working through one's emotions actually helps to ensure objectivity, rather than hinder it. Moreover, despite being emotionally burdening, the respondents mostly evaluated working with distressing data as (at least partially) positive experience (88%), since it enabled them to better understand societal issues, academic work and last but not least provided an opportunity for personal growth.

Risk factors. Our data confirms Hypothesis 1.3 stating that CMC linguistic features influenced annotators' and researchers' evaluation of the comment, since this was confirmed by more than half of the respondents. They reported to have been the most influenced by incorrect language use and orthographic specifics, such as the use of all caps, unusual punctuation and emoticons or emojis. Usually, they perceived these elements as underlining the comment's unacceptability, or the use of these features led them to inadvertently negatively judge such a comment or its author. While not a characteristic of CMC as such, some noted that, at the first sight, their perception of longer and linguistically correct comments was biased towards perceiving them as acceptable despite their possibly unacceptable content. This is important to take into account in any future study of SUD which should address the question how these potentially influencing linguistic factors should be considered in the annotation process and analysis.

We cannot confirm Hypothesis 1.4 stating that longer immersion in the data results in stronger distress for the annotators, since our data shows conflicting results. A third of the respondents reported a link between their working pace and the intensity of emotions experienced, however the resulting reaction is not unanimous. A slightly larger group represents those who reacted strongly at the beginning of the annotation process or in shorter working sessions, and experienced numbness in longer sessions, while a slightly smaller number of respondents said their experience was the contrary (longer sessions provoked hyper-sensitivity). We should also note that, despite possible drawbacks of continuous immersion, a certain level of immersion is necessary in order to ensure annotation consistency. This was also pointed out by a good third of respondents who said that with time they became more confident in their classification, as well as more efficient and attentive. In addition, researchers stated that they found qualitative analysis more emotionally demanding, but that quantitative methods also produced some emotional stress during interpretation when actual content had to be linked to the raw numbers.

We cannot confirm Hypothesis 1.5 stating that personal circumstances increase the intensity of emotional response to the data. Although almost half of the participants reported that they have a personal circumstance that could influence their reaction and work (political and sexual orientation, religious belief, gender, expertise), none of them, in a separate question, reported developing an intense emotional reaction to the data. Furthermore, few of those who reported experiencing a traumatic event before or during the work process, did not report experiencing an intense emotional reaction to the data. This is corroborated by respondents' answers to a direct question about their perceived link between their life experience and their work on the project, since only 15% indicated that their personal circumstances might have impacted their emotional reaction. A higher share of respondents (60%), however, stated that their life experience might have impacted their mental reaction to the comments by developing a more critical/sensitive appreciation of the comment or higher concern for objectivity. A small share of respondents (15%) completely rejected any possible impact of their personal experience on their project work. This corroborates the claim of Moran and Asquith (2020) who suggest that the potential hazard is greater for those with a personal connection to the topic, but that this same connection can provide the researchers with crucial skills in conducting analysis and developing resilience against negative consequences. We therefore assume that the indicated personal circumstances might not have extensively influenced respondents' emotional and mental reactions since they were aware of them and actively worked to prevent any subjectivity resulting from these circumstances.

4.2. Mitigation measures

Preparatory phase. Most of the respondents (60%) said that guidelines and initial training provided sufficient support to feel emotionally and professionally well-prepared for their work, however over a third (36%) felt that the preparatory phase could be improved. The main criticism was pointed towards guidelines which were, according to the respondents, at least in the first stages of the annotation campaign, too basic (i.e. did not contain enough difficult illustrative examples). Furthermore, some doubted that the short preparatory training (1/2-day) could equip the annotators with the knowledge required to recognize normalized SUD, while others felt very insecure in their role of annotators fearing to assume the role of censors. Therefore, although with certain restraint, we reject *Hypothesis 2.1* stating that initial training and annotation guidelines do not ensure adequate support to preserve the well-being of collaborators. However, as will

become clear in the next section, training and guidelines are only one type of supportive elements, and as such do not provide a comprehensive support (i.e. mental and emotional) needed by the researchers.

Execution phase. During the project, the project team organised regular meetings, opened a Google group for discussions and provided regular feedback from an expert annotator on comments with the highest inter-annotator disagreement. Group meetings and online discussions were not intensively attended, while the feedback from the expert annotator was in general appreciated. Among the most frequent coping mechanisms, the respondents reported taking a break, talking to somebody or reflecting on their emotions. Most respondents (60%) claimed they felt the need to discuss their work experience with others, among which they listed family members and partners (60%), friends (60%) and fellow annotators (30%), while some also talked to the campaign leader, a psychologist and groups who were targets of SUD comments.[4] Therefore, we confirm *Hypothesis 2.2* stating that most annotators and researchers seek additional support outside the project. However, literature as well as some respondents' remarks (e.g., wishing to talk to someone acquainted with the task) show that peer support should not be relied upon as the main/only source of support.

Debriefing. Although only one respondent directly expressed the wish to have an organized debriefing at the end of the annotation campaign, several authors (Fahie, 2014; Goodrum & Keys, 2007) indicate that debriefings are important when working with distressing data since such data can cause re-living of negative experiences (as was reportedly the case for one annotator in our project as well). Due to several annotation campaigns that run in succession and many annotators participating in more than one campaign, we did not think of providing separate debriefings after each campaign. Professional debriefing was thus only offered after the penultimate campaign, but was not delivered due to COVID-19 limitations and low interest among the annotators. Our survey included a question regarding the offer for a debriefing, and 40% of respondents indicated interest in such a session because they wanted to exchange experience with other annotators and learn about possible impacts of SUD of which they were not aware. On the other hand, the majority of those who deemed debriefing as unnecessary argued that there is no need since they are not this sensitive or because the emotions they had experienced did not influence other areas of their life.

In sum, our experience shows that the supporting activities offered during the project (meetings, discussion group) were not extensively used, despite the clear general interest/need expressed in the survey for exactly such activities (i.e. more meetings, creating a chat room for all annotators, group debriefing, consultations with psychologist). Therefore, we assume that (1) this was caused by inadequate preparation of researchers by the project leader (failing to explain the possible hazards); and (2) that the project leader did not encourage the collaborators actively enough to use the supportive measures. The reflection that this study enabled has also revealed a lack of institutional support with regard to ensuring the well-being of researchers, since Ethics Committees mainly highlight the responsibilities of researchers towards research participants and much less for themselves and other team members.[5]

5. Conclusion

In this paper, we looked at the emotional hazard that arises from working with possibly distressing linguistic data. The data was obtained through a self-observation survey filled in by 25 annotators and linguistic researchers of socially unacceptable Facebook comments. The study shows that annotation and research of SUD is perceived as distressing work which impacts researchers' activities on several levels: from the need to secure their well-being to the need to adopt techniques which can ensure output quality similar to other (non-distressing) research projects (e.g. avoiding common pitfalls such as lack of objectivity or frequent difficulties with concentration). It is clear from this study that emotional hazard is not bound to field work alone, but is also present in less obvious settings, such as desktop linguistic annotation and research. Although the study is based on introspection and a relatively small sample, the trends in the results as well as the growing body of related work on the topic leave no doubt of the importance of the issues linked to working with distressing data. In order to ensure high quality research outputs and continuous well-being of the entire research team, it is, therefore, important to conduct a comprehensive risk assessment, not only for research participants but for all collaborators, and to propose mitigation measures which cover the entire research cycle.

6. Acknowledgements

We thank the anonymous reviewers for constructive comments. This paper was funded by the Slovenian Research Agency within the projects »Resources, methods, and tools for the understanding, identification, and classification of various forms of socially unacceptable discourse in the information society« (J7-8280, 2017–2019), »Linguistic landscape of hate speech on social media« (N06-0099, 2019–2023) and the research programme "Slovene language – basic, contrastive and applied studies" (P6-0215).

7. References

Asquith, N. L. (2010). Verbal and textual hostility in context. In N. Chakraborti (Ed.), *Hate Crime: Concepts, policy, future directions* (pp. 99–123). Willan.

[4] The sum is greater than 100 because some indicated more than one support group of individuals.

[5] An anonymous reviewer also suggested that it is good practice to include into the annotation protocol the possibility for the collaborators to reject parts of the annotation/research task if perceived too disturbing.

Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, 61(8), 1139–1160.

Bride, B. E., Robinson, M. M., Yegidis, B., & Figley, C. R. (2004). Development and validation of the secondary traumatic stress scale. *Research on Social Work Practice,* 14(1), 27–35.

Delgado, R. (2019). *Understanding words that wound.* Routledge.

Dickson-Swift, V., James, E. L., Kippen, S., & Liamputtong, P. (2008). Risk to researchers in qualitative research on sensitive topics: Issues and strategies. *Qualitative Health Research*, 18(1), 133–144.

Dickson-Swift, V., James, E. L., Kippen, S., & Liamputtong, P. (2009). Researching sensitive topics: Qualitative research as emotion work. *Qualitative Research*, 9(1), 61–79.

Fahie, D. (2014). Doing sensitive research sensitively: Ethical and methodological issues in researching workplace bullying. *International Journal of Qualitative Methods*, 13(1), 19–36.

Figley, C. R. (1995). *Compassion fatigue: Coping with secondary traumatic stress disorder in those who treat the traumatized.* Routledge.

Fincham, B., Scourfield, J., & Langer, S. (2008). The impact of working with disturbing secondary data: Reading suicide files in a coroner's office. *Qualitative Health Research*, 18(6), 853–862.

Goodrum, S., & Keys, J. L. (2007). Reflections on two studies of emotionally sensitive topics: Bereavement from murder and abortion. *International Journal of Social Research Methodology*, 10(4), 249–258.

Grant, A. (2018). Shock and offence online: The role of emotion in participant absent research. In T. Loughran & D. Mannay (Eds.), *Emotion and the Researcher: Sites, Subjectivities, and Relationships* (Vol. 16). Emerald Publishing Limited.

Hanna, E. (2019). The emotional labour of researching sensitive topics online: Considerations and implications. *Qualitative Research*, 19(5), 524–539.

Jackson, S., Backett-Milburn, K., & Newall, E. (2013). Researching distressing topics: Emotional reflexivity and emotional labor in the secondary analysis of children and young people's narratives of abuse. *Sage Open,* 3(2), 2158244013490705.

Ljubešić, N., Fišer, D., & Erjavec, T. (2019). The FRENK Datasets of Socially Unacceptable Discourse in Slovene and English. In K. Ekštein (Ed.), *Text, Speech, and Dialogue: Proceedings / 22nd International Conference, TSD 2019, Ljubljana, Slovenia, September 11–13, 2019* (pp. 103–114). Springer. https://doi.org/10.1007/978-3-030-27947-9

McCosker, H., Barnard, A., & Gerber, R. (2001). Undertaking sensitive research: Issues and strategies for meeting the safety needs of all participants. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 2(1).

Moran, R. J., & Asquith, N. L. (2020). Understanding the vicarious trauma and emotional labour of criminological research. *Methodological Innovations,* 13(2), 2059799120926085.

Stern, S. R. (2003). Encountering distressing information in online research: A consideration of legal and ethical responsibilities. New Media & Society, 5(2), 249–266.

Vehovar, V., Povž, B., Fišer, D., Ljubešić, N., Šulc, A., & Jontes, D. (2020). Družbeno nesprejemljivi diskurz na facebookovih straneh novičarskih portalov. *Teorija in Praksa*, 57(2), 622–645.

Familiarisers and Taboo Vocatives in Twitter Discourse

Ignacio M. Palacios Martínez
University of Santiago
E-mail: ignacio.palacios@usc.es

Abstract

Nominal vocatives that serve as terms of address have been the object of extensive study in a wide array of offline communication contexts (Braun, 1988; Leech, 1999; MacCarthy and O' Keefe, 2003; Rendle-Short, 2010; Heyd, 2014; Clancy, 2015; Adams, 2018; Kluge and Moyna 2019). Yet, this is not the case in CMC, more particularly in Twitter-based discourse, where research thus far has been far more limited. While some such studies (Hastrdlová 2009) have been concerned with general address terms in English-language IRC, or how people address one another in emails, weblogs, online chats and mobile messages (Anglemark 2009, Asprey and Tagg 2019), others have focused on address terms in the educational domain (Belz and Kinginger 2002, Nguyen and Kellogg, 2005).

This paper examines two groups of vocatives in Twitter-based discourse: familiarisers (man, brother, dude) and taboo address terms (bastard, dickhead). To this end, a sample of 1,500 posts are analysed, these having been extracted from the accounts of three well-known London rappers between July 2018 and May 2021.

The findings show the high frequency of these terms of address in the exchanges between these singers and their followers. A large majority convey positive meanings and, unexpectedly, only six (e.g. bitch, wanker) can be regarded as offensive. Brother is by far the most common, followed by mate, man, guys and pals. Some are preceded by you and my, and may be grammaticalized, as in the case of man, which can function as a regular noun (Cheshire 2013). Although they are found mainly in final position, vocatives in initial position seem to be more frequent here than in spontaneous speech. Furthermore, whereas in offline communication such vocatives serve to convey a variety of pragmatic meanings, especially those associated with the marking of interpersonal relationships between the participants in conversation, this function is especially prevalent in Twitter-based discourse, where such address terms help to reinforce the role of participants as members of a group and of a community of practice, thus contributing to the creation of the sense of affiliation which plays such an important role in microblogging (Zappavigna 2012).

Keywords: address terms, vocatives, Twitter

References

Adams, Z. (2018). "I don't know why man's calling me family all of a sudden": Address and reference terms in grime music. *Language & Communication* 60, pp. 11–27.

Anglemark, L. (2009). Address Terms in Computer Mediated Communication: Email, Chat and Weblogs. Ph.D dissertation. Uppsala University.

Asprey, E.; Tagg, C. (2019). The pragmatic use of vocatives in private one-to-one digital communication. *Internet Pragmatics* 2(1), pp. 83-111.

Belz, J.A.; Kinginger, C. (2002). The cross-linguistic development of address form use in telecollaborative language learning: Two case studies. *The Canadian Modern Language Review/La Revue Canadienne des langues vivantes* 59(2), pp. 189-214.

Braun, F. (1988). *Terms of Address*. Berlin, New York, Amsterdam: Mouton de Gruyter.

Cheshire, J. (2013). Grammaticalisation in social context: The emergence of a new English pronoun. *Journal of Sociolinguistics* 17 (5), pp. 608-633.

Clancy, B. (2015). "Hurry up baby son all the boys is finished their breakfast." Examining the use of vocatives as pragmatic markers in Irish traveller and settled family discourse. In C. P. Amador-Moreno, K. McCafferty and E. Vaughan (Eds.), *Pragmatic Markers in Irish English*, Amsterdam: Benjamins, 229-247.

Hastrlová, Š. (2009). Language of Internet Relay Chat (Reconsidering (Im)politeness in the Context of IRC). Ph.D dissertation. University of Masarykova at Brno.

Kluge, B.; Moyna, M.I. (Eds.) 2019. *New Perspectives on Address Research*. Benjamins: Amsterdam.

Leech, G. (1999). The distribution and function of vocatives in American and British English. In H. Hasselgård and S. Oksefjell (Eds.). *Out of Corpora: Studies in Honour of Stig Johansson*. Amsterdam: Rodopi, 107-120.

McCarthy, M.; A. O'Keefe. (2003). "What's in a name?" Vocatives in casual conversation and radio phone-in calls. In P. Leistna and C. Meyer (Eds.), *Corpus Analysis: Language Structure and Language Use*. Amsterdam: Rodopi, 153-185.

Ngugen, H.T.; Kellogg, G. (2005). Emergent identities in on-line discussions for second language learning. *The Canadian Modern Language Review/La Revue Canadienne des langues vivantes* 62(1), pp. 111-136.

Rendle-Short, J. 2010. 'Mate' as a term of address in ordinary interaction. *Journal of Pragmatics* 42, pp. 1201-1218.

Zappavigna, M. (Ed.). 2012. *Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web.* New York: Continuum.

Parallels Between Spoken and CMC Language: Do Tweets Reflect Spoken Language Choices?

Adriana Picoral[1], Elisa Stumpf[2], Larissa Goulart[3], Isabella Calafate de Barros[1], Bruna Sommer-Farias[4], Marine Laísa Matte[5], Marina Carcamo Garcia[1], Mariana Centanin Bertho[1]

[1]University of Arizona, [2]Universidade Federal de Pelotas, [3]Northern Arizona University, [4]Michigan State University, [5]Universidade Federal do Rio Grande do Sul

E-mail: adriananaps@email.arizona.edu, elisa.stumpf@gmail.com, lg845@nau.edu, isabellacb@email.arizona.edu, fariasbr@msu.edu, marina.laisa@gmail.com, mcarcamogarcia@email.arizona.edu, marianabertho@email.arizona.edu

Abstract

While a lot is known about pronominal variation in spoken Brazilian Portuguese produced by different speech communities across the country, little is known about how language variation is realized in computer-mediated communication (CMC). This study investigates whether the same type of pronominal variation can be found in Brazilian Portuguese tweets. Our logistic regression results for two variables, first person plural subject pronoun (i.e., nós vs. a gente, 'we') and second person singular subject pronoun (i.e., você vs. tu, 'you'), show that Twitter users present the same variation across geographic regions found in spoken data, making the case for more language variation studies using CMC data.

Keywords: pronominal variation, language variation and change, Twitter language use

1. Introduction

The field of language variation and change (LVC) has focused its efforts on understanding language use through the collection of large-scale, random sampling of spoken language production from different communities. It has been often the case, however, that corpora built for LVC studies have not always been large-scale and not entirely randomly sampled (Tagliamonte, 2012). On the other hand, CMC data (e.g., tweets and other social media posts) allow for large-scale and randomly sampled corpora, which can be extremely valuable for the study of language variation. The question of whether CMC data can be used to reveal patterns of language variation, serving as a proxy for spoken usage, has been explored in other contexts (see for example De Decker, Vandekerckhove, and Sandra, 2016, for a study on Flemish and Dutch; and Willis, 2020, on Welsh). Linguistic variation studies with social media language have been conducted, but these have focused mainly on English and Spanish variation (Friginal, Waugh, & Titak, 2018; Willis, 2020). The present study aims at bridging this gap by shedding light on pronominal variation in Brazilian Portuguese on Twitter data, with the goal of comparing the CMC language produced in this social media platform to findings on spoken Brazilian Portuguese.

2. Pronominal variation in Brazilian Portuguese

Pronominal variation in spoken Portuguese has been extensively studied by the field of sociolinguistics (see for example Vianna, & dos Santos Lopes, 2012 for first person pronoun studies and Scherre, Andrade & Catão, 2021 for second person pronoun). Innovative pronouns have been introduced to Brazilian Portuguese through grammaticalization of nouns and other pronouns, with the function of individualizing or disambiguating the referent (Vianna, & dos Santos Lopes, 2012). Originating from *vossa mercê* (i.e., your mercy) (Do Monte, 2019), a common form of address used in Brazil when it was still a Portuguese colony, the pronoun *você* (you) alternates with *tu* (you). Introduced through a similar process of grammaticalization, the replacement of the *nós* pronoun (i.e., we) by the innovative pronoun *a gente* (literal translation meaning "the people") has been an ongoing process of change in Brazilian Portuguese (Zilles, 2005).

The choice between *tu* vs. *você* (second person singular pronouns) and *nós* vs. *a gente* (first person plural pronouns) as subject pronouns in modern Brazilian Portuguese is determined by linguistic (e.g., verb tense) and extra-linguistic factors (e.g., geographic region) (Carvalho, Freitas, & Favacho, 2020). As such, these variables are useful for the study of CMC language.

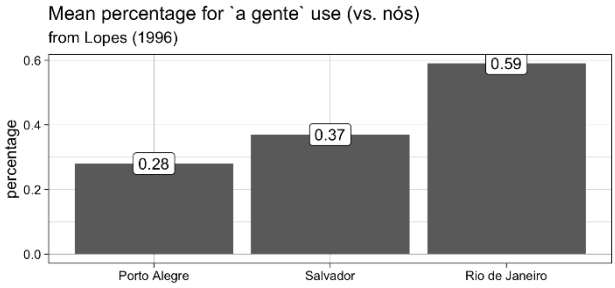


Figure 1: Percentages for *a gente* use for first person plural subject pronoun (vs. *nós*) across metropolitan regions in Brazil according to Lopes (1996)

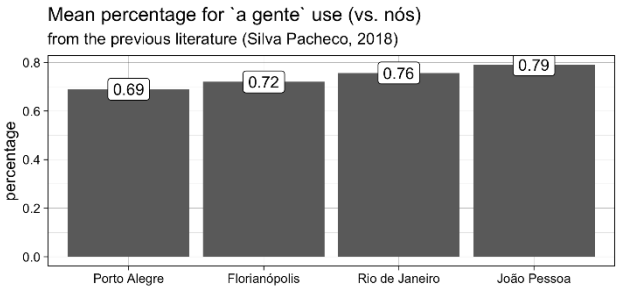


Figure 2: Percentages for *a gente* use for first person plural subject pronoun (vs. *nós*) across metropolitan regions in Brazil according to previous literature (Silva Pacheco, 2018)

The previous literature on subject pronoun variation in Brazilian Portuguese has shown that the choice of second person pronouns (*tu* vs. *você*) is strongly determined by dialect (Guimarães, & de Sousa Pereira, 2019), while the choice between *a gente* vs. *nós* might not be as easily determined by region. The pronoun *tu* is more frequently used in the south of Brazil, including the regions of Porto Alegre and Florianópolis, while *você* is preferred in São Paulo, Rio de Janeiro and Pernambuco (Scherre, Andrade, & Catão, 2020; Scherre, Yacovenco, & de Paiva, 2019). Research on the first person pronoun, on the other hand, shows that there is overall a stronger preference for *a gente* across Brazil (Silva Pacheco, 2018), but as it can be seen in Figures 1 and 2, regional differences are harder to determine. Figure 1 shows the percentages of *a gente* use for three regions in Brazil (Porto Alegre, Salvador, and Rio de Janeiro) according to the previous literature, indicating Porto Alegre is the region with the lowest incidence of *a gente* (Lopes, 1996). Figure 2 shows the percentage of *a gente* use for Porto Alegre, Florianópolis, Rio de Janeiro and João Pessoa, with all regions displaying an incidence of *a gente* (compared to *nós*) of .69 to .79 (Silva Pacheco, 2018).

3. Data and methods

3.1. Our corpus

Our first step was to compile a corpus of tweets that contained the pronoun forms of interest, i.e., tweets that either had *nós* or *a gente* for first person plural subject pronouns and that contained *tu* or *você* for second person singular pronouns (and all orthographic variants for each pronoun, such as *a gnt* for *a gente* and *vc* for *você*). To that end, between June 20 and July 10, 2020, we searched and downloaded 10,413 tweets through the Twitter streaming API in R (R Core Team, 2020) using the rtweet package (Kearney, 2019). Multistage sampling was used, with tweets being randomly selected from specific geolocations. While tweets provide a large volume of data to be analyzed, they do not contain metadata related to extra-linguistic variables such as age and social class. The geographic metadata, however, is a great advantage of using social media language (Eisenstein, 2018; Grieve et al, 2019). Based on previous literature on age, race and gender of Twitter users (see for example Huang et al., 2020), we can assume that our corpus represents a younger (mid 20s on average) and whiter (50%) population compared to the general demographics in Brazil. For region specification, we used geolocation codes (i.e., latitude and longitude) for six major metropolitan areas in Brazil (i.e., Salvador, Recife, João Pessoa, Rio de Janeiro, São Paulo, Florianópolis, and Porto Alegre). This was a necessary step to control for location, since the location entered by Twitter users in their profiles is not reliable.

3.2. Envelope of variation

Our linguistic variables are the alternation between two subject pronouns in Brazilian Portuguese for two different pronoun types: the choice between *a gente* and *nós* for first person plural subject pronoun, and the choice between *tu* and *você* for second person subject pronoun.

After a set of random tweets containing subject pronoun forms was collected, the next step was to code each tweet as belonging or not in our envelope of variation (Milroy, & Gordon, 2008). When determining the envelope of variation, it is important to identify the contexts that 1) allow for variation, and 2) are not categorically encoded with one of the variants (Tagliamonte, 2012). We decided to keep only subject pronouns in our study, since object pronouns present different types of variation depending on the context. For example, after the preposition *com* (i.e., with), different pronoun variations are possible: with *tu*, a contraction is used, forming *contigo* (example 1); with *nós*, a contraction is used as well, forming *conosco* (example 2); with the pronouns *você* (example 3) and *a gente* (example 4) contraction with preposition is never realized. These variations go beyond the scope of this study.

tô deixando essa bomba **contigo**
(I) am leaving this bomb **with+you**
'I'm leaving this problem to you'
Você estará sempre **conosco**
You be+future always **with+us**
'You will always be **with us'**
Ontem eu tive esse sonho, nele encontrava **com você**
Yesterday I had this dream, in+it (I) met **with you**
'Yesterday I had this dream, in which I ran into you'
ou a mente acaba **com a gente**
or the mind ends **with us**
'or our mental health destroys us'

Another challenge when retrieving tweets with the target pronoun is that it is impossible to rely on form alone to decide whether the token is a pronoun inside our envelope of variation. Examples (5-6) show how the same form can be a subject pronoun (5a-5b), an object pronoun (6a), or a preposition plus article contraction (6b-6c).

tokens included in the analysis:
só que **nos** vivemos ataques das gringas por mais de um mês
only that **we** lived attacks from foreigners for more than a month
'It's just that we've lived under foreign attack for more

than a month'
vou ai qualquer hora pra **nos** toma uma gelada
(I) go there any time for **us** drink a cold
'I'll be there anytime for us to drink a beer'
tokens eliminated from the analysis:
infelizmente esse direito **nos** foi tomado
unfortunately this right **us** was taken
'unfortunately this right was taken from us'
E tu **nos** meus sonhos
And you **in+the** my dreams
'And you in my dreams'
Mas **nós** últimos dias não consigo nem me ajudar
But **in+the** last days not can not even me help
'But lately I can't help even myself'

For this reason, automatic parsing of these tweets for our specific purposes (i.e., retrieving subject pronouns) was not accurate enough. We opted then for hand coding each tweet to ensure variation in the data reflected actual language variation. About 30% of tweets initially collected were eliminated from our analysis for not being in our envelope of variation. We also eliminated tweets that were song lyrics and quoted text, since these do not represent speaker choice. Table 1 shows the number of tokens (i.e., first person plural and second person singular subject pronoun use instances) across seven regions in Brazil.

	1st person plural	2nd person singular
Florianópolis	355	144
João Pessoa	419	247
Porto Alegre	215	606
Recife	397	236
Rio de Janeiro	554	634
Salvador	118	409
São Paulo	420	477
TOTAL	2,478	2,753

Table 1: Token count for first person plural and second person singular tokens across regions in Brazil.

3.3. Factor groups

Since we are using Twitter data, our extra-linguistic factor is region only. We built our corpus from random tweets in certain regions, meaning our Twitter users are not known to us, and as such we are unable to obtain other factors related to their background, such as age, gender, and socio-economic status. Our goal was to build a representative corpus of Brazilian Portuguese on Twitter, which reflects the choice of subject pronouns. Thus, we opted to obtain data from a large number of users (N=1,679 for first person and N=1,999 for second person), instead of focusing only on known users (e.g., public figures) for which we would have detailed demographic information. While other variables were coded (e.g., use of internet language, generic vs. specific referent), we discuss in this paper pronominal variation across geographical regions only.

3.4. Data analysis

Linguistic data tend to be unevenly distributed even in controlled experimental conditions and even more so in more naturalistic settings of data collection (e.g., sociolinguistic interviews, and social media data), which renders statistical tests such as ANOVA not as reliable as originally intended (Sankoff, & Labov, 1979). The solution proposed by scholars in the field of Language Variation and Change (LVC) is estimation of maximum likelihood, or logistic regression, with sum contrasts (Sankoff, 1982) and a binary dependent variable representing the discrete language choice (Grieve, 2012). We report on two models, one for the first person plural subject pronoun choice (i.e., *nós* vs. *a gente*) and another for the second person singular subject pronoun choice (i.e., *tu* vs. *você*). We include in our plots not only the probability estimates across different regions, but also the confidence intervals. All modeling was run in R with the stats package (R Core Team, 2020).

4. Results

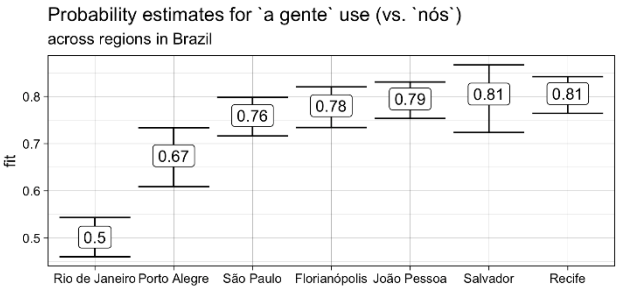


Figure 3: Probability estimates for *a gente* use for first person plural subject pronoun (vs. *nós*) across metropolitan regions in Brazil

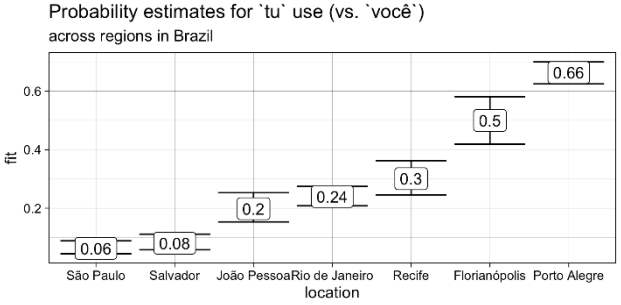


Figure 4: Probability estimates for *tu* use for second person singular subject pronoun (vs. *você*) across metropolitan regions in Brazil

	Our Results	Silva Pacheco, 2018
João Pessoa	.79	.79
Florianópolis	.78	.72
Porto Alegre	.67	.69
Rio de Janeiro	.50	.76

Table 2: Comparison of our current results for first person plural pronoun variation, showing probability fit for *a gente* use across regions compared to average percentage use per region reported in Silva Pacheco, 2018.

Figure 3 shows the probability estimates output by the first person plural logistic regression model. Although the entire factor group is significant to predict *a gente* use in the data ($p < .001$), Rio de Janeiro is the only region that displays lower probability for *a gente* (at .50), with all other regions displaying higher probabilities (between .67 and .81). Our results closely match the most recent literature on first person subject pronoun variation for different regions, with the exception of Rio de Janeiro, as shown in Table 2. A number of factors may explain the permanence of *nós* in Rio de Janeiro, including the use of third singular agreement (e.g. "nós vai" - we *goes*) to resolve temporal ambiguities. Indeed 70% of *nós*/we tokens are combined with a third singular agreement verb form in Rio de Janeiro. This combination of the standard first plural pronoun + a third singular verb form might serve as an index of local identity, which could be due to the less monitored discourse found in Twitter in comparison to other written data (McCulloch, 2019), or even a mockery strategy. A further discussion on this is beyond the scope of this paper and will be explored in future research.

Regarding the results for second person singular subject pronoun choice, Figure 4 shows the probability estimates output by the logistic regression model with *tu* as the reference variant. Location is a significant predictor in the second person pronoun model ($p < .001$). These findings match previous language variation results for this variable (Scherre, Andrade, & Catão, 2020). Table 3 shows our results compared to percent averages of *tu* use found in previous studies (Scherre, Andrade, & Catão, 2020; Scherre, Yacovenco, & de Paiva, 2019). As shown, the percentages of *tu* use for Porto Alegre and Florianópolis seem to be overall lower in our Twitter corpus compared to spoken language. This can be explained by the stigma associated with *tu* across most of Brazil, as evidenced in the use of this pronoun to characterize the speech of low-class people in soap operas (Scherre, Yacovenco, & de Paiva, 2019). Due to Twitter's more public discourse, users from Southern Brazil might be adapting their language to a broader audience. However, the order in which regions use *tu* remains mostly the same. Porto Alegre is the region that favors *tu* at a higher probability of .66. Florianópolis

	Our Results	Scherre et al. 2021; 2019
Porto Alegre	.66	.91
Florianópolis	.50	.76
Rio de Janeiro	.24	.27
Recife	.30	.14
João Pessoa	.20	.04
Salvador	.08	.01
São Paulo	.06	.00

Table 3: Comparison of our current results for second person singular, showing probability fit for *tu* (you) use across regions compared to average percentage use per region according to previous literature (Scherre, Andrade, & Catão, 2020; Scherre, Yacovenco, & de Paiva, 2019).

is significantly different from Porto Alegre and São Paulo/Rio de Janeiro, showing a 50/50 preference for *tu*. Salvador and São Paulo are the regions that least favor the use of *tu,* with probability estimates for this variant of .08 and .06 respectively.

5. Discussion

In this paper, we present regional variation of pronominal choice in Brazilian Portuguese tweets, with the purpose of verifying parallels between CMC and spoken language choices. Overall, our results corroborate findings from previous studies, showing that language used in CMC reflects the speech community patterns. In the case of first person plural pronouns, Twitter data shows probabilities of *a gente* use very similar to those found in the previous literature for Porto Alegre, Florianópolis and João Pessoa (Silva Pacheco, 2018). However, the data also shed new light on the regional conditioning of this variable by showing that, different from other metropolitan regions analyzed, *nós* is highly retained in Rio de Janeiro. The reasons for the permanence of *nós* in this location in particular will be discussed in future research.

For second person singular pronouns, our results match the previous literature, showing a larger probability of *tu* use in the south of Brazil and *você* preference in other regions (Scherre, Andrade & Catão, 2020). In our data, we also noticed the frequent use of *vc*, an emerging pronoun form in CMC, which does not exist in oral production. This shortened form of *você* will be discussed in comparison to *tu* and *você* in future research.

Twitter is a valuable source of data for the study of language variation. With that being said, we need to be careful about how we go about using social media data. With this study, we had to opt for hand coding tweets since the accuracy of current NLP tools available for Portuguese did not allow for the automatic extraction of our variable

at an acceptable accuracy rate. Without careful coding of tweets, to ensure we included pronoun forms that were inside our envelope of variation, the comparison between CMC language production and spoken language across different regions might have been invalid.

6. Future direction

Our next step is to use our hand-annotated corpus of tweets as ground truth for the automatic parsing of the data for subject pronoun variation study purposes. This will enable us to develop an automated method for extraction of language data from CMC that allows us to scale our investigation and analyze a larger number of factors that affect language choice.

7. References

Carvalho, H., Freitas, M., & Favacho, L. (2020). A variação dos pronomes sujeitos nós e a gente: a fala culta de Fortaleza em cena. *Revista (Con)Textos Linguísticos*, *14*(27), 30-45.

De Decker, B., Vandekerckhove, R., & Sandra, D. (2016). When two basic principles clash: about the validity of written chat language as a research tool for spoken language variation. Flemish chatspeak as a test case. *Journal of Language Contact*, *9*(1), 101-129.

Do Monte, V. M. (2019). Eu lhe alembro a você: sobre o lugar de vossa mercê e você na história do português. *Working Papers em Linguística*, *20*(2), 85-108.

Eisenstein, J. E. (2018). Identifying Regional Dialects in On-Line Social Media. In C. Boberg, J. Nerbonne, & D. Watt (Eds.), *The Handbook of Dialectology* (pp. 368–383). Wiley.

Friginal E., Waugh O., Titak A. (2018) Linguistic variation in Facebook and Twitter posts. In: E. Friginal (Ed) *Studies in Corpus-Based Sociolinguistics* (pp. 342–63). New York: Routledge.

Grieve, J. (2012). Sociolinguistics: Quantitative Methods. In C.A. Chapelle (Ed.). *The Encyclopedia of Applied Linguistics* (pp. 100-150). Oxford. https://doi.org/10.1002/9781405198431.wbeal1090

Grieve, J., Montgomery, C., Nini, A., Murakami, A., & Guo, D. (2019). Mapping lexical dialect variation in British English using Twitter. *Frontiers in Artificial Intelligence*, *2*(11), 1-18.

Guimarães, T. D. A. A. S., de Araújo, A. A., & de Sousa Pereira, M. L. (2019). Os pronomes você (s) e cê (s) no falar de Fortaleza-CE. *Letras Escreve*, *8*(2), 75-101.

Huang, X., Xing, L., Dernoncourt, F., & Paul, M. J. (2020). Multilingual Twitter corpus and baselines for evaluating demographic bias in hate speech recognition. *LREC 2020*.

Lopes, C. R. D. S. (1996). Nós por a gente: uma contribuição da pesquisa sociolingüística ao ensino. *Diversidade Lingüística e Ensino*, 115-123.

McCulloch, G. (2019). *Because internet: Understanding how language is changing*. Random House.

Kearney, (2019). rtweet: Collecting and analyzing Twitter data. *Journal of Open Source Software*, *4*(42), 1829. https://doi.org/10.21105/.

Milroy, L., & Gordon, M. (2008). *Sociolinguistics: Method and interpretation* (Vol. 13). John Wiley & Sons.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Sankoff, D. (1982). Sociolinguistic method and linguistic theory. In S*tudies in Logic and the Foundations of Mathematics* (Vol. 104, pp. 677-689). Elsevier.

Sankoff, D., & Labov, W. (1979). On the uses of variable rules. *Language in Society*, 8(2), 189-222.

Silva Pacheco, C. D. (2018). Diachrony and synchrony of first-person plural pronouns Nos and A Gente in Brazilian Portuguese and in Uruguayan Portuguese. *Revista de Estudos da Linguagem*, *26*(1), 221-253.

Scherre, M. M. P., Andrade, C. Q. & Catão, R. C. (2021). Por onde transitam o tu e o voice no nordeste? The whereabouts of personal pronouns tu and você ('you') in the Brazilian Northeast. *Revista de Letras*, *40*(1), 1-40.

Scherre, M. M. P., Andrade, C. Q. & Catão, R. C. (2020). Redesenhando o mapa dos pronomes tu/você/cê/ocê no português brasileiro falado In: *Resumos expandidos das Comunicações Individuais do V Congresso Nacional de Estudos Linguísticos - CONEL*, 2019, Vitória- ES.

Scherre, M. M. P., Yacovenco, L. C., & de Paiva, M. D. C. (2019). An overview of Luso-Brazilian sociolinguistics: Second person pronouns. In D. A. Koike & D. Tejedo-Herrero (Eds.) *Lusophone, Galician, and Hispanic Linguistics* (pp. 246-262). Routledge.

Tagliamonte, S. A. (2012). *Variationist sociolinguistics: Change, observation, interpretation* (Vol. 39). John Wiley & Sons.

Thomé-Williams, A. C. (2004). Sociolinguistic aspects of forms of address in Portugal and Brazil: tu or você. *Intercultural Communication Studies*, *13*(3), 85-99.

Vianna, J. B. S., & dos Santos Lopes, C. R. (2012). A competição entre «nós» e «a gente» nas funções de complemento e adjunto: desvendando outras portas de entrada para o pronome inovador. *Caligrama: Revista de Estudos Românicos*, 17(2), 137-161.

Willis, D. (2020). Using social-media data to investigate morphosyntactic variation and dialect syntax in a lesser-used language: Two case studies from Welsh. *Glossa: a journal of general linguistics*, 5(1).

Zilles, A. M. (2005). The development of a new pronoun: The linguistic and social embedding of a gente in Brazilian Portuguese. *Language Variation and Change*, 17(1), 19-5.

Annotation of Attack Speech Acts in Wikipedia Talk Pages

Céline Poudat
Université Côte d'Azur, CNRS, BCL, France
E-mail: celine.poudat@univ-cotedazur.fr

Abstract

Wikipedia talks are a particular CMC genre in which editors discuss the progress of the encyclopedic articles and settle their editorial conflicts. Our study concentrates on the linguistic expression of conflicts in Wikipedia. A pragmatic annotation task was carried out on the Wikiconflits corpus (Poudat et al., 2017, available on https://repository.ortolang.fr/api/content/comere/v2/cmr-wikiconflits.html) which was originally developed in the CoMeRe project (Chanier et al., 2014). Two talk pages were added to the Wikiconflits corpus: the talk pages on the articles *Vladimir Poutine* (324 posts) and *Féminisme* (423 posts).

Each talk page is divided into threads, corresponding to the page sections. A systematic categorisation task of the discussion threads of the corpus allowed us to distinguish between conflictual and non-conflictual threads (Poudat and Ho-Dac 2019).

In order to observe the linguistic expression of conflicts, we first concentrated on agreement and disagreement (Poudat&Ho-Dac, 2019; Poudat 2018) and we led a systematic annotation task of the disagreements in the Wikiconflits corpus and also marked up explicit agreements in contrast (Poudat 2018).

In a second phase, we concentrated on the development of a system of attack speech acts in Wikipedia talks, on which we will concentrate in the present poster.

Our objective is twofold: on the one hand, we will present the system of speech acts we developed and on the other hand, we will report the manual annotation task we carried out, which is currently being completed by a crowdsourcing experience in order to complete and assess the annotations.

Keywords: Wikipedia talk pages, conflicts, corpus pragmatics, annotation

References

Chanier, T., Poudat, C., Sagot, B., Antoniadis, G., Wigham, C., Hriba, L., Longhi, J. et Seddah, D. (2014). The CoMeRe corpus for French: structuring and annotating heterogeneous CMC genres. In Beißwenger, M., Oostdijk, N., Storrer, A. et van den Heuvel, H. (ed.), Special Issue of *Journal of Language Technology and Computational Linguistics* "Building and annotating corpora of computer-mediated discourse: Issues and Challenges at the Interface of Corpus and Computational Linguistics", pp. 1-30.

Poudat, Céline, Grabar, Natalia, Paloque-Berges, Camille, Chanier, Thierry et Kun, Jin (2017). « Wikiconflits : un corpus de discussions éditoriales conflictuelles du Wikipédia francophone » *in* Wigham, C.R & Ledegen, G., *Corpus de communication médiée par les réseaux : construction, structuration, analyse*. Collection Humanités numériques. Paris : L'Harmattan, pp. 19-36.

Poudat, C. et Ho-Dac, L.-M. (2019). Désaccords et conflits dans le Wikipédia francophone. In Col, G. et Hanote, S. « Accord et désaccord », *Travaux linguistiques du Cerlico*, n°29, pp. 155-176.

Poudat, C. (2018). Explorer les désaccords dans les fils de discussion du Wikipédia francophone. In Iezzi D. F., Celardo, L. et Misuraca, M., *Actes des JADT 2018*, Rome, vol. 2, pp. 602-610.

Social Media Data as a Naturalistic Test Bed for Studying Sociolinguistic and Psycholinguistic Patterns in Verb Spelling Errors

Hanne Surkyn, Reinhild Vandekerckhove, Dominiek Sandra
CLiPS Research Center, University of Antwerp
E-mail: hanne.surkyn@uantwerpen.be, reinhild.vandekerckhove@uantwerpen.be, dominiek.sandra@uantwerpen.be

Abstract

In the present paper, we review the benefits and challenges when using social media data to examine sociolinguistic and psycholinguistic patterns in the production of spelling errors on Dutch verb homophones. Unlike most previous research on verb homophones, which is often of an experimental nature, we examined teenagers' private, informal chat conversations (see Surkyn et al., 2019, 2020, 2021). We found both new psycholinguistic patterns as well as patterns of homophone intrusions that were previously attested in an experimental context (see e.g., Sandra, Frisson, & Daems, 1999). Moreover, our research adds a social dimension to this primarily psycholinguistic field of research by testing the effect of gender, age and educational track. We argue that these social media data form an ideal testing ground for validating results from psycholinguistic experiments outside the lab *and* for revealing new insights into the social and cognitive processes underlying the spelling of Dutch verb homophones.

Keywords: social media writing, spelling errors, verb homophones, sociolinguistics, psycholinguistics

1. Introduction

The goal of the present paper is to show how corpus research based on social media data can complement experimental approaches to questions in psycholinguistics. Besides emphasizing the benefits, we also consider potential pitfalls.

In our previous papers (Surkyn, Vandekerckhove, & Sandra, 2019; Surkyn, Vandekerckhove, & Sandra, 2020; Surkyn, Vandekerckhove, & Sandra, 2021) our major goal was to use spelling errors on social media as another window on spellers' cognitive infrastructure. The importance of such an interdisciplinary approach was previously emphasized by Gries (2012, 2013) and Meyer et al. (2018), among others. In his articles, Gries (2012, 2013) explicitly pleads for a combination of (sociolinguistic) corpus research and (psycholinguistic) experimental research. Meyer et al. (2018), who studied response latencies in conversations, argue that corpus analyses of natural language data and psycholinguistic experiments can be complementary. They showed that a combination of both methods is needed in order to get a better understanding of the cognitive processes in conversations (Meyer et al., 2018, p. 9).

In our studies, we focused on the effect of psycholinguistic and sociolinguistic factors on the production of verb spelling errors. Regarding the psycholinguistic component, we looked for (a) the effect of homophone dominance (see below) that was found in experimental research of, for example, Sandra and colleagues (1999) (Surkyn et al., 2020), and (b) for analogical effects (previously attested in reading experiments by Ernestus & Mak, 2005) on the spelling of partially homophonous verb forms (Surkyn et al., 2021). Hence, we tested the ecological validity from previous experimental findings via corpus research. In recent studies, Schmitz, Chamalaun, and Ernestus (2018) and Gahl and Plag (2019) also investigated spelling errors using social media corpora. However, they used (public) tweets, whereas our data consist of informal, private chat conversations.

In addition, we used social media data to uncover social patterns. We examined the effect of the social factors gender, age and educational track on the number of verb spelling errors chatters produced. The impact of these social factors on the production of verb spelling errors in social media data had not been investigated before. As we combined both a sociolinguistic and psycholinguistic perspective, we were also able to test whether the social and mental factors interact.

The added value of the present paper is that it combines results and insights from several case studies of ours that have not been brought together hitherto, thus offering a bird's-eye view on how corpus research on social media data may contribute to a better understanding of an extensively researched and multifaceted topic in Dutch psycholinguistic research.

2. The Spelling of Dutch Verb Forms

In our studies (Surkyn et al., 2019; Surkyn et al., 2020, Surkyn et al., 2021), we focused on the production of spelling errors on (partially) homophonous Dutch verb forms. The spelling of regular verb forms in Dutch reflects the application of two important spelling principles: a phonological and a morphological one. As a result of the phonological principle, we spell words the way they sound. However, in some cases, this principle is overruled by the morphological principle, which guarantees that morphemes are spelled in the same way across all forms in which they occur (e.g., *rond*: adjective, 'round' – *de ronde*: noun, 'the round' – *ronden*: verb, 'to round'). This is indeed the case for homophonous verb forms, which are pronounced in the same way but spelled differently, such as *(ik) word* ('I become') and *(hij) wordt* ('he becomes'), both pronounced as [wɔrt]. The stem *word*, which is also the spelling of the 1st person singular present tense, is consistently spelled with a <d> (in accordance with the morphological principle) although a [t] is heard in the forms presented here, because word-final obstruents in Dutch are devoiced. In the spelling of the 2nd and 3rd person singular present tense, the letter *-t* is added to the stem (*word* + *-t*), as a [t] is heard in verb forms with the same grammatical function (e.g., *hij spreekt*, 'he speaks'), which betrays the presence of a suffix. Despite these descriptively simple rules, spelling errors on Dutch verb homophones are extremely persistent and occur in both experienced and inexperienced writers.

By means of speeded dictation tasks, researchers revealed the cognitive mechanisms behind these persistent verb

spelling errors (e.g., Frisson & Sandra, 2002; Sandra, Frisson, & Daems, 1999). As spellers of Dutch are seldom confronted with these homophonous verb forms (only in about 5 to 10% of all verb types and tokens, see Verhaert, Danckaert, & Sandra, 2016), the (conscious) application of the spelling rules has not become an automatic process. Consequently, spelling errors often occur when there is insufficient time for writers to apply the morphographic spelling rules in working memory. In these cases of working memory overload, the higher frequency spelling, which is most strongly anchored in the mental lexicon and is thus more accessible, may pop up. Hence, most errors occur on the lower-frequency homophone. This effect of homophone intrusions is known as the homophone dominance effect and reflects an interplay between conscious processes in working memory and unconscious processes in the mental lexicon (Sandra et al., 1999). The effect of homophone intrusions has also been observed in other languages such as French (e.g., Largy, Fayol, & Lemaire, 1996).

3. Methodology

3.1. Corpus

Unlike most previous research on verb spelling errors, in which researchers used psycholinguistic spelling experiments (see e.g., Frisson & Sandra, 2002; Sandra et al., 1999; Verhaert et al., 2016), we studied these errors in a natural writing context, i.e., in the informal social media writing of Flemish teenagers. Our unique corpus consists of more than 400,000 private posts produced by Flemish adolescents, aged between 15 and 20 years, on Facebook Messenger and WhatsApp, two popular applications for instant messaging.[1]

The data collection was supported by teachers and school principals. Pupils (and for minors also their parents) were asked permission to store and linguistically analyze their anonymized messages. For all participants we have meta-information on their gender, educational track and age. Table 1 shows the distribution of the participants and target verb forms over these social determinants.

3.2. Data Processing

In our studies, the dependent variable was the spelling performance on regular Dutch verb forms. The spelling of the <d>, <t>, or <dt> ending of these verb forms was manually encoded as 'correct' or 'incorrect' in order to determine the error risk. In automated prefiltering processes, all potentially relevant verb forms were extracted from the corpus by selecting all tokens ending in <d>, <t>, or <dt>. Next, all false positives were removed manually. We analyzed the binary response variable (correct/incorrect) by performing generalized linear mixed-effects models, using the glmer function from the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in the R statistical software package (R Core Team, 2014). We opted for a stepwise forward procedure, first adding the random factors chatters and lemma to the models, followed by the fixed factors (i.e., the sociolinguistic and psycholinguistic variables) and their interactions. Likelihood ratio tests ($\alpha = .05$) were used in order to decide whether adding an extra factor to the models accounted for significantly more variance in the data. Finally, Variance Inflation Factors (VIFs) were calculated in order to check whether there was no collinearity among the predictors.

Variable	Variable Level	Partici-pants	Target Forms
Gender	Boys	631	4,907
	Girls	657	9,838
Educational Track	General Education	559	4,646
	Technical Education	370	6,939
	Professional Education	359	3,160
Age	Younger (15-16)	978[2]	7,396
	Older (17-20)	897	7,349

Table 1: Distribution of the data over gender, educational track and age.

4. Results

4.1. Confirmatory analyses

Our results confirmed the conclusions from experimental psycholinguistic research. We attested the effect of homophone dominance (Surkyn et al., 2020), previously demonstrated in speeded dictation tasks by Sandra and colleagues (1999), in two different types of homophonous verb forms, i.e., stem-final <d> verbs (e.g., *word-wordt*, 'become-becomes') and prefixed verbs (e.g., *verandert-veranderd*, 'changes-changed').[3] Hence, we showed that the experimental results from psycholinguistic research generalize to real-life writing situations. Moreover, we found no interaction between this (psycholinguistic) pattern of homophone intrusions and the social factors: The effect of homophone dominance was equally strong in all social groups (Surkyn et al., 2020). In other words, spellers from all social groups fell prey to the same type of homophone intrusions when the mental lexicon took over from successful rule application.

The latter finding, too, aligns with the model of Sandra and colleagues (1999) on the nature of the cognitive processes underlying the persistence of these spelling errors (see

[1] The original corpus, as composed by Lisa Hilte and described in Hilte (2019, p. 17), also included messages of 13-and 14-year-olds. We excluded their messages from our analyses because we have much less data (and fewer target forms) from these youngest teenagers than from the older teenagers.

[2] Note that the sum of the number of younger and older adolescents does not correspond to the total number of participants. This is due to the fact that the same participants can occur several times, at different ages, in the corpus if they submitted both recent and older messages.

[3] Both homophone types were misspelled in 29% of the cases (see Surkyn et al., 2020 for more details about the error rates).

section 1. Introduction). Their account predicts that spellers with different sociodemographic profiles differ only at the level of conscious processes in working memory, not at the level of unconscious processes, i.e., the automatic retrieval of orthographic representations. Hence, factors affecting the error rates (such as gender and age) should not influence the error pattern (i.e., the effect of homophone dominance). This prediction was thus confirmed via our corpus research.

4.2. Novel findings

4.2.1. Sociolinguistic patterns

Our corpus studies also revealed new psycholinguistic patterns and sociolinguistic patterns that had not been attested for these spelling errors. At the sociolinguistic level, our results showed that girls, students in more theory-oriented educational tracks and older teenagers produce significantly fewer verb spelling errors than boys, students in more practice-oriented educational tracks and younger teenagers (Surkyn et al., 2020; Surkyn et al., 2021). These effects were observed for both verb homophones and partially homophonous verb forms. The effect of gender may be explained by a difference in spelling attitude, more particularly by the more norm-compliant behavior of women (Labov, 2001, p. 293) and their tendency to avoid stigmatized forms (Tagliamonte, 2011, p. 32).

We related the effect of educational track to a difference in rule knowledge. Teenagers in the more theory-oriented tracks are likely to have a better knowledge of the verb spelling rules, as their official curriculum focuses more on spelling than the curricula for more practice-oriented tracks (VVKSO, 2006, 2014). This, in turn, might lead to a more outspoken attitude towards error avoidance in more theory-oriented tracks. Moreover, teenagers of different educational tracks might differ in terms of their exposure to the Dutch language in general and to written texts in particular (Surkyn et al., 2021). Previous research of Hilte, Vandekerckhove, and Daelemans (2018) showed that Dutch is much more often not the home language of the students in more practice-oriented tracks than of the students in more theory-oriented tracks, at least for adolescents in this corpus. Moreover, students in more theory-oriented tracks are likely to have more experience with, for example, novels and proof-read texts, i.e., with standard writing. This might have a positive effect on the quality of their orthographic representations (The Lexical Quality Hypothesis, e.g., Perfetti & Hart, 2002, and Perfetti, 2007). Research of Rahmanian and Kuperman (2019), for instance, demonstrated that exposure to common spelling errors leads to more uncertainty and less stable orthographic representations.

We attributed the effect of age to both a difference in attitudes and a difference in rule knowledge. On the one hand, younger teenagers can be expected to attach less importance to standard writing norms and correct spelling. In previous sociolinguistic studies, researchers found that non-conformist language behavior of teenagers increases until their mid-teens and decreases as they approach adulthood (e.g., Holmes, 1992, p. 184; Coates, 1993, p. 94). On the other hand, older students probably have more experience with formal standard writing (through books and proof-read texts, for example) than younger ones. Furthermore, they might have a better command of the verb spelling rules. In a recent study, Chamalaun, Bosman, and Ernestus (2021) found that older adolescents had a better grammatical knowledge and also made fewer verb spelling errors.

4.2.2. Psycholinguistic patterns

With regard to the cognitive factors, we found two types of analogical effects on the spelling of partially homophonous past participles spelled with a <d> ending (e.g., *genoemd*, 'named', Surkyn et al., 2021). These past participles are only partially homophonous with other forms in the inflectional paradigm (e.g., with the 3rd person singular present tense form *noemt*, 'names').[4] This type of verb forms had not been experimentally investigated before and could now be analyzed on the basis of our extensive corpus of social media data. Firstly, our results revealed that the spelling of these past participles was affected by intraparadigmatic support for the correct <d> spelling: Stronger <d> support (relative to <t> support) in the verb's inflectional paradigm caused significantly fewer (incorrect) <t> spellings. Thus, the frequency of morphologically related forms, i.e., other forms in the verb's inflectional paradigm, had an effect on the spelling of the partially homophonous past participles. This analogical effect of intraparadigmatic <d> support, which we found in a natural writing context, is comparable with the intraparadigmatic analogical effects attested in speeded reading experiments targeting inflected verb forms in Dutch (Ernestus & Mak, 2005).

Secondly, our study showed an effect of support for the correct spelling from the bigram straddling the morpheme boundary (e.g., <md> in *gedroomd*, *ge-* + *droom* + *-d*, 'dreamed'). The more frequent a final bigram ending in <d> (e.g., <md>) occurred in inflected forms across all verbs in Dutch relative to the verb-final bigram ending in <t> (e.g., <mt>) (i.e., the higher the value of this ratio), the smaller the error rates (Surkyn et al., 2021). Interestingly, the frequency of the word-final bigrams in words of lexical categories other than verbs had no significant impact on the error rates. This indicates that this is not a purely orthographic effect, as the <d> and <t> spellings reflect phonologically conditioned allomorphs of the past participle suffix. Hence, our study revealed both an analogical effect at the intraparadigmatic level and an analogical effect at the sublexical level. This effect of bigram <d> support, an analogical effect at the sublexical level, was not previously observed in experimental research. This emphasizes the importance of combining experimental and non-experimental research.

[4] The overall error rate of these past participles was 11% (see Surkyn et al., 2021 for more details about the error rates).

5. General Discussion

Working with social media data offers several advantages. Firstly, researchers of social media can gain new insights into sociolinguistic patterns in a psycholinguistic research domain (e.g., spelling errors). Our studies demonstrated, for instance, that girls, while falling prey to the same type of errors as boys, attach more importance to the correct spelling of verb forms than boys, even in the informal setting of private social media conversations where norm-adherence generally is much less of an issue that in other writing contexts. In addition, analyses of social media data can reveal both experimentally attested and novel psycholinguistic patterns, which may form the starting point of new psycholinguistic experiments. In our study (Surkyn et al., 2021), an effect of intraparadigmatic and bigram support for the correct spelling of partially homophonous past participles was established for the first time. In future research, these analogical effects could be tested by means of spelling experiments. Vice versa, psycholinguistic researchers can indicate which research questions can also be investigated through corpus studies.

Furthermore, CMC data offer the opportunity to test the ecological validity of psycholinguistic models in a naturalistic writing situation. Based on our corpus, we were able to check whether the psycholinguistic effect of homophone dominance (previously attested in spelling experiments) could be generalized to the natural context of informal social media writing. At first glance, it may seem odd to compare spelling errors produced in formal (psycholinguistic) spelling experiments with those produced in informal social media conversations. Indeed, comparing the number of verb spelling errors in both contexts would be problematic, as strict adherence to standard writing norms in informal CMC is generally not expected and pursued. Perhaps one might also expect a radically different distribution of error types. However, our results showed that in both formal (experimental) and informal (social media) contexts, the same type of verb spelling errors occurred for homophonous verb forms (i.e., intrusions of the higher-frequency form). Hence, our results corroborate the relevance of comparing spelling patterns attested in psycholinguistic experiments with those in informal social media data.

We could also test a crucial prediction of the psycholinguistic model, i.e., that the social and cognitive factors should not interact. Furthermore, results from large dataset analyses have a larger statistical power than findings in a typical psycholinguistic experiment, which often involves relatively few subjects (e.g., twenty to thirty participants). Our use of a large social media corpus is in line with the trend in psycholinguistics over (roughly) the past ten years to use larger data sets (often referred to as 'megastudies') to arrive at new insights (e.g., Keuleers & Balota, 2015).

However, working with CMC corpora like ours also poses a few challenges. In contrast to psycholinguistic experiments, in which an equal distribution of items and subjects can easily be controlled, this type of CMC datasets is often unbalanced in terms of subjects and items. In our studies, for example, some chatters (mostly girls and students in technical education) were overrepresented in terms of number of posts, tokens, and target forms (see Surkyn et al., 2020, pp. 430–431). This imbalance was a consequence of the data collection method. Adolescents voluntarily donated their CMC data and were free to determine how many conversations they submitted. As a result, some chatters sent much more data than others (see Hilte 2019, pp. 9–11). However, extra analyses revealed that the presence of chatters that contributed lots of data or very little data had no significant impact on the results of our mixed model analyses (see Surkyn et al., 2021).

In terms of items, some verb forms occurred highly frequently in the corpus, others only once or twice. This is in fact predicted by Zipf's law (Zipf, 1949), which states that there are big frequency differences among the words in a language. Our studies showed that these distributional differences should be handled with caution, as a couple of high-frequency verbs (e.g., *vinden* 'to find', *bedoelen* 'to mean', *zeggen* 'to say') had a great impact on our results (e.g., Surkyn et al., 2020, pp. 447–448). This forced us to use extra techniques, such as resampling and frequency cutoffs, in order to limit the impact of outliers. The interquartile method for the removal of outliers could not solve our problem of skewed data, as this method drastically reduced the size of our dataset and, hence, the number of target forms. Thus, multi-disciplinary approaches require researchers to be careful to avoid statistical pitfalls.

Sometimes, the outliers were informative in their own right. More particularly, they revealed the implicit learning of incorrect forms. In the dataset of partially homophonous past participles ending in <d>, the verb form *gezegd* ('said') stood out. This past participle had an extremely high occurrence frequency compared to the other forms (i.e., 24% of all targets). At the same time, it was surprisingly often misspelled as *gezegt* (a non-existent spelling), i.e., in 27% of the cases. We argued that chatters, being very often exposed to this incorrect spelling, might have implicitly learned this form. This, in turn, might have resulted in the unconscious 'imitation' or even storage of a competing orthographic representation (Surkyn et al., 2021). Thus, this misspelling (*gezegt*) may have become a typical marker of the informal online writing of some chatters. Hence, one can study unintended effects of implicit learning (of e.g., misspellings) in the context of CMC research.

The informal, private nature of our corpus also has its drawbacks at the level of data sharing: Although anonymized, the corpus cannot be made publicly available because this is not allowed by the ethical committee of our university for the sake of privacy protection. A chat corpus like ours obviously consists of highly personal, private conversations, mostly between (close) friends that sometimes share sensitive information. While names (of people, places, events…) and other personal data (e.g., telephone numbers) have been erased, perfect anonymization presents huge challenges for this type of data.

We can conclude that social media data are highly useful for studying old and new sociolinguistic and psycholinguistic patterns and for the 'ecological validation' of psycholinguistic findings. Furthermore, these naturalistic data prove to be suitable for examining the interaction between social and mental processes. At the same time, it is important not to lose sight of the potentially unbalanced nature of the dataset, which requires extra caution at the methodological level.

6. References

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, pp. 1-48.

Chamalaun, R., Bosman, A., & Ernestus, M. (2021). The Role of Grammar in Spelling Homophonous Regular Verbs. *Written Language & Literacy*, 24(1), pp. 38-80.

Coates, J. (1993). *Woman, men and language. A sociolinguistic account of sex differences in language*. London: Longman.

Ernestus, M., & Mak, W. M. (2005). Analogical effects in reading Dutch verb forms. *Memory & Cognition*, 33, pp. 1160-1173.

Frisson, S., & Sandra, D. (2002). Homophonic Forms of Regularly Inflected Verbs Have Their Own Orthographic Representations: A Developmental Perspective on Spelling Errors. *Brain and Language*, 81, pp. 545-554.

Gahl, S., & Plag, I. (2019). Spelling errors in English derivational suffixes reflect morphological boundary strength: A case study. *The Mental Lexicon*, 14, pp. 1-36.

Gries, S. T. (2012). Corpus linguistics, theoretical linguistics, and cognitive/psycholinguistics: Towards more and more fruitful exchanges. Corpus Linguistics and Variation in English, *Language and Computers*, 75, pp. 41-63.

Gries, S.T. (2013). Sources of variability relevant to the cognitive sociolinguist, and corpus- as well as psycholinguistic methods and notions to handle them. *Journal of Pragmatics*, 52, pp. 5-16.

Hilte, L. (2019). *The social in social media writing: The impact of age, gender and social class indicators on adolescents' informal online writing practices*. Antwerp: University of Antwerp (doctoral thesis).

Hilte, L., Vandekerckhove, R., & Daelemans, W. (2018). Adolescents' social background and non-standard writing in online communication. *Dutch Journal of Applied Linguistics*, 7, pp. 2-25.

Holmes, J. (1992). *An introduction to sociolinguistics*. London / New York: Longman.

Keuleers, E., & Balota, D. A. (2015). Megastudies, crowdsourcing, and large datasets in psycholinguistics: An overview of recent developments. *The Quarterly Journal of Experimental Psychology*, 68, pp. 1457-1468.

Labov, W. (2001). *Principles of Linguistic Change*, Vol. 2: Social Factors. Malden, MA: Blackwell Publishers

Largy, P., Fayol, M., & Lemaire, P. (1996). The homophone effect in written French: The case of verb-noun inflection errors. *Language and Cognitive Processes*, 11, pp. 217-255.

Meyer, A. S., Alday, P. M., Decuyper, C., & Knudsen, B. (2018). Working Together: Contributions of Corpus Analyses and Experimental Psycholinguistics to Understanding Conversation. *Frontiers in Psychology*, 9, pp. 525.

Perfetti, C. A., (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific studies of reading*, 11, pp. 357-383.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, 11, pp. 67-86.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienne, Austria.

Rahmanian, S., & Kuperman, V. (2019). Spelling errors impede recognition of correctly spelled word forms. *Scientific Studies of Reading*, 23, pp. 24-36.

Sandra, D., Frisson, S., & Daems, F. (1999). Why simple verb forms can be so difficult to spell: the influence of homophone frequency and distance in Dutch. *Brain and Language*, 68, pp. 277-283.

Schmitz, T., Chamalaun, R., & Ernestus, M. (2018). The Dutch verb-spelling paradox in social media. A corpus study. *Linguistics in the Netherlands*, pp. 111-124.

Surkyn, H., Vandekerckhove, R., & Sandra, D. (2019). Errors Outside the Lab: The Interaction of a Psycholinguistic and a Sociolinguistic Variable in the Production of Verb Spelling Errors in Informal Computer-Mediated Communication. In J. Longhi & C. Marinica (Eds.), *Proceedings of the 7th Conference on CMC and Social Media Corpora for the Humanities (CMC-Corpora2019)* (pp. 59-62). Paris.

Surkyn, H., Vandekerckhove, R., & Sandra, D. (2020). From experiment to real-life data: social factors determine the rate of spelling errors on rule-governed verb homophones but not the size of the homophone dominance effect. *The Mental Lexicon*, 15(3), pp. 422-463.

Surkyn, H., Vandekerckhove, R., & Sandra, D. (2021). The impact of analogical effects and social factors on the spelling of partially homophonous verb forms in informal social media writing. *Written Language & Literacy*, 24(1), pp. 1-37.

Tagliamonte, S. (2011). *Variationist Sociolinguistics: Change, observation, interpretation*. Oxford: Wiley-Blackwell.

Verhaert, N., Danckaert, E., & Sandra, D. (2016). The dual role of homophone dominance. Why homophone intrusions on regular verb forms so often go unnoticed. *The Mental Lexicon*, 11, pp. 1-25.

VVKSO. (2006). *Leerplan Nederlands secundair onderwijs bso*. [Curriculum Dutch secondary education bso]. Brussel: Vlaams Verbond van het Katholiek Secundair Onderwijs. Retrieved from: http://ond.vvkso-ict.com/leerplannen/doc/Nederlands-2006-020.pdf

VVKSO. (2014). *Leerplan Nederlands secundair onderwijs aso-kso-tso*. [Curriculum Dutch secondary education aso-kso-tso]. Brussel: Vlaams Verbond van het Katholiek Secundair Onderwijs. Retrieved from: http://ond.vvkso-ict.com/leerplannen/doc/Nederlands-2014-001.pdf

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley Press.

Buzz or Change: How the Social Network Structure Conditions the Fate of Lexical Innovations on Twitter

Louise Tarrade[1], Jean-Pierre Chevrot[2], Jean-Philippe Magué[1]
Ens de Lyon[1], Université Grenoble Alpes[2]
E-mail: louise.tarrade@ens-lyon.fr, jean-pierre.chevrot@univ-grenoble-alpes.fr, jean-philippe.mague@ens-lyon.fr

Abstract

The diffusion process of linguistic innovations has long been a topic of interest in sociolinguistics (Weinreich et al., 1968) and many studies have highlighted the influence of social structures on change (Labov, 2001; Milroy & Milroy, 1997). The recent access to massive social network data and the advent of computational sociolinguistics (Nguyen et al., 2016) allow an approach to this phenomenon that combines a large amount of data and a fine-grained temporality.

Using methods from both computational sociolinguistics and network science, we focus on the diffusion of lexical innovations and we ask what differentiates, after an expansion phase, those that stabilize within our observation period from those that are eventually abandoned. In particular, we examine the impact of the social structure of linguistic communities on these diffusion and acceptance processes.

We rely on a corpus of French tweets, that spans from 2012 to 2019 and includes about 600 million tweets from more than two million users. Based on the evolution over time of the rate of use of each linguistic form, we select those that appear during the period covered by the corpus and then we distinguish the forms that stabilize from those that eventually die out. By modeling the trajectories, we then identify the three characteristic periods of the diffusion of an innovation (Fagyal et al., 2010). Finally, by establishing the network of contact between users on the basis of their followers and followees, we examine the circulation of forms between them at different periods, and identify factors that condition the stabilization or not of innovations.

This poster will present the methodologies used to identify linguistic innovations and to model their trajectory. We will also present the first results on the connection between the evolution of forms and the structure of the contact network.

Keywords: computational sociolinguistics, lexical change, diffusion of innovation, network structure, Twitter data

References

Fagyal, Zsuzsanna, Samarth Swarup, Anna María Escobar, Les Gasser, and Kiran Lakkaraju. 2010. "Centers and Peripheries: Network Roles in Language Change." *Lingua* 120 (8): 2061–79. https://doi.org/10.1016/j.lingua.2010.02.001.

Labov, William. 2001. *Principles of Linguistic Change Vol. 2 : Social Factors*. Oxford: Blackwell. Milroy, James, and Lesley Milroy. 1997. "Network Structure and Linguistic Change." In *Sociolinguistics*, edited by Nikolas Coupland and Adam Jaworski, 199–211. London: Macmillan Education UK. https://doi.org/10.1007/978-1-349-25582-5_17.

Nguyen, Dong, A. Seza Doğruöz, Carolyn P. Rosé, and Franciska de Jong. 2016. "Computational Sociolinguistics: A Survey." *Computational Linguistics* 42 (3): 537–93. https://doi.org/10.1162/COLI_a_00258.

Weinreich, Uriel, William Labov, and Marvin I Herzog. 1968. "Empirical Foundations for a Theory of Language Change." *WP Lehmann-Y. Malkiel (Hrsgg.), Directions for Historical Linguistics, Austin/London*.

Webcare in Building Virtual Rapport: Apologising in Hotel Management Responses to Negative Online Reviews

Ly Wen Taw[1, 2], Shamala Paramasivam[2], Alan Libert[1], Christo Moskovsky[1]
CLiPS Research Center, [1]*University of Newcastle, NSW, Australia*
[2]*Universiti Putra Malaysia, Selangor, Malaysia*
E-mail: taw.lywen@uon.edu.au, shamala@upm.edu.my, alan.libert@newcastle.edu.au

Abstract

With technological advancement, social media platforms are empowering consumers' voices in conveying their feedback regarding products or services digitally. This online feedback mechanism has become an influential electronic Word-of-Mouth (eWOM) that has an impact on consumer purchase decisions. Consequently, online service recovery is essential to not only restore customer satisfaction, but also to manage online reputation management via eWOM. Therefore, showing webcare, which refers to the engagement in online interactions with consumers, is crucial in various industries, including the hotel industry. This digital engagement can build trust and rapport with customers, leading to customer loyalty and higher business revenues. Service recovery in digital platforms has been given considerable attention in marketing or business research, and it has increasingly gained more attention in recent years in the field of linguistics. However, service recovery from a linguistic perspective, particularly in Southeast Asia, has received scant attention in the literature. Using Spencer-Oatey's (2008) Rapport Management Model, this study investigates the speech act of apology in hotel responses on TripAdvisor used by Malaysian five-, four- and three-star hotels in dealing with complaints in negative online reviews. This study utilised the research approach of Herring's (2004) Computer-Mediated Discourse Analysis (CMDA) in data analysis. The findings identified nine apology strategies. Among the use of strategies, the five-star hotel managements strikingly displayed a preference for the strategy of offer of repair and recorded the lowest frequencies for the strategy of explanation of situation in comparison to the four- and three-star hotel managements. This study not only provides cultural insights into Malaysian managerial practices in rapport management strategy in service recovery, but also offers practical implications for service managers in the hospitality industry about maximising online service recovery efforts to achieve optimum customer satisfaction.

Keywords: apology, webcare, online reviews, eWOM, hotel responses, rapport

References

Herring, S. C. (2004). Computer-Mediated Discourse Analysis: An Approach to Researching Online Behaviour. In S. A. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for Virtual Communities in the Service of Learning* (pp. 338-376). New York: Cambridge University Press.

Spencer-Oatey, H. (2008). Culturally Speaking: Culture, Communication and Politeness Theory. London: Continuum.

Recontextualization Strategies in Medical YouTube Videos

Anna Tereszkiewicz, Magdalena Szczyrbak
Jagiellonian University
E-mail: anna.tereszkiewicz@uj.edu.pl, magdalena.szczyrbak@uj.edu.pl

Abstract

The poster presents the results of a study focusing on medical knowledge popularisation videos shared on the most popular YouTube channel in Poland (Najprościej mówiąc). The research question behind the study was: What rhetorical strategies are used in the videos to recontextualise medical knowledge in the multimodal and participatory context of YouTube? The analysis was based on Luzón's (2013, 2015) classification of strategies used to recontextualise scientific knowledge in scientific weblogs, and Myers' (2010) discussion of engagement strategies in blogs. The initial classifications were revised and expanded to suit the audio-visual context and include the practices employed by the YouTubers.

The analysis shows that the authors resort to a diversified set of recontextualisation strategies, fully exploiting the affordances offered by the multimodal character of the medium. The video makers tailor medical information and engage the audience by using explanation, exemplification, comparison, references to shared knowledge, by the use of informality and slang, as well as humour and intertextuality. Specialist knowledge is recontextualised by means of fictionalised scenes, role-play, demonstrations, as well as animations. The use of music, headlines and diverse visuals is also exploited as a means of educating and entertaining the audience. Quite interestingly, certain glocalisation patterns were found in the strategies, with the YouTubers utilising references to the Polish socio-cultural context, which may not necessarily be clear to a foreign viewer. The poster presents the respective practices identified in the videos, with supporting examples and an interpretation of the findings. In focusing on and revealing the practices used by Polish YouTubers, the study contributes to the growing body of research into spoken online discourse across different linguistic backgrounds. The results of the study may be of interest to researchers investigating popularisation videos in an increasingly multilingual and multicultural setting of YouTube and other broadcasting platforms.

Keywords: medical discourse, YouTube, recontextualisation, science communication

References

Luzón, M.J. 2013. Public communication of science in blogs: Recontextualizing scientific discourse for a diversified audience. *Written Communication* 30 (4): 428-457.

Luzón, M.J. 2015. Recontextualising expert discourse in weblogs: Strategies to communicate health research to experts and the interested public. In: Gotti, M., Maci, S.M., Sala, M. (eds), *Insights into Medical Communication*. Bern: Peter Lang, 331-351.

Myers, G. 2010. The Discourse of Blogs and Wikis. London: Continuum.

Look Back Without Anger: Recapitulation of the Corpus *What's up, Switzerland?*

Simone Ueberwasser
University of Zurich
Email: simone.ueberwasser@uzh.ch

Abstract

In 2014, a team of Swiss linguists and computational linguists started a collection of WhatsApp messages in all four national languages. Their methodology was partly based on an earlier collection of text messages (SMS) within the same team, other parts of their approach were newly designed. Looking back on the now finished project, many of these steps were fruitful while others could have been approached in a better way in hindsight. In the course of the last year, the team received several questions from other linguists intending to perform a collection of their own to know more about the methodology applied by the Swiss team. In this presentation, the important steps of the collection will be recapitulated and evaluated in order to give others some insight into advantages and disadvantages of the steps taken by the Swiss team.

Keywords: WhatsApp, corpora, best practice

1. Motivation

In 2009, a team of Swiss linguists and computational linguists collected some 24,000 text messages (Dürscheid and Stark, 2011) based on a methodology developed in Belgium (Cougnon, 2008). The corpus is freely available for non-commercial use (Stark et al., 2015) The bi-national project based on this corpus lead to four dissertations and received a lot of attention, both from the Swiss press and from the scientific community. Accordingly, the team decided to put together a follow-up project based on WhatsApp messages. While the SMS had been collected with assistance from the major Swiss provider of mobile services, the new WhatsApp collection of 2014 had to be attacked without that support. However, other strategies such as the set-up of a questionnaire and some of the data processing or anonymization were taken over from the preceding project. This second collection also resulted in a bi-national research project with five doctoral students.

2. Steps of the collection

In a call in the press, Swiss citizens were asked to send one or more backup files created within WhatsApp to a specified e-mail address. The project website hosted a detailed instructions in all four national languages plus English, such as to explain exactly how to generate these backup files both on Android and iOS. The verbal instruction was underlined by graphical elements such as printscreens from either operating system or by charts to explain the workflows.

When a backup file was received by the inbox, a script was triggered to perform the following steps:

- Save the backup file as a text file with a unique identifier key as a name
- Reply to the e-mail address with a request to:
  - Approve the use of the sent in date on the project website
  - Fill in an online questionnaire with 16 questions
  - Send an e-mail to all communication partners in that specific chat asking for their permission to use their messages as well.

In order to identify the chat, the ID given to the backup file by the script was used in this communication process, while communication partners were identified by the nicknames they had given themselves in WhatsApp. This step lead to several types of chats (Ueberwasser and Stark, 2017, p. 112). There were those, where all messages in a specific chat could be used for research, because all communication partners had given their consent. In an ideal case, at least some demographic information was available for all communication partners, too. For many chats, however, consent was not given by all communication partners and accordingly demographic information was missing, too. Messages with consent asked for less data processing in the next step than those without.

3. Preparing the corpus

The first and foremost processing step was to ensure that only messages with consent were used in the corpus. Since the communication would have looked strange if only one person's text had been available, those without context were replace with codes like *redactedQ12tokens55characters* to show that the disguised chat originally contained 12 tokens making up a total of 55 characters. This and all other steps taken are documented in more detail in the corpus documentation (Stark and Ueberwasser, 2020). The next was to anonymize information such as names, addresses and numbers. Here, the methodology developed for the SMS corpus was followed and the lists of information created in that project were used.

Since one of the postdocs in the project was to work on emojis, some time and thought was given to the question how emojis can be queried. On the one hand, the browsing tool has to support the representation of the Unicode characters for emojis, on the other hand, it has to be possible to search for e.g. the character representing a specific emoji. The biggest challenge, however, is to find groups of emojis like "all smiling faces" or "all cat faces". In order to enable this kind of query, it was decided to store the messages in two different layers where the first layer represented the original message while in the second layer, the emoji was replaced by the official description of the emoji as used by the Unicode Consortium, e.g. *emojiQsmilingCatFaceWithOpenMouth*, where *emojiQ* is the internal encoding to ensure that this kind of encryption is not mixed

up with written out text.
Within each chat, there is a chance of finding technical messages written by WhatsApp itself, like ”XY is leaving the chat” or ”XY changed their name”. In a multilingual corpus, these messages appear in the language of the person who sent the chat in. In order to ensure that they are not mistaken for texts written by communication partners, they were standardized and put in a form of e.g. *action-QuserOUT*.
The team also realized that they received more than one copy of the same chat when both or even several communication partners sent in the same chat. In that situation, the latest (and thus longest) version was kept while all others were deleted.

4. Annotation of languages

As mentioned before, the corpus contains data in all four national language plus sporadic texts in other languages such as Spanish or Bosnian. One would think that annotating languages should be easy these days as with all the tools that are available. However, more than halve of the data is in Swiss German, a dialect with little resemblance to Standard German and without any spelling standards. Some computational tools are available e.g. for the standardisation of Swiss German, however at the time of the creation of the corpus, nothing could be found that would have allowed to annotate parts of text in Swiss German within a corpus in other languages. It was thus decided to use student assistants instead. They were to read the first 250 messages in each chat and then decide the main language of this chat. This information was used to create sub-corpora per language. The aim was to get quick and rather reliable result such as to allow the doctoral students to start quickly with their theses in the respective languages.
Later on in the project, the team from Leipzig developed a tool to differentiate between Standard and Swiss German. To a great part, the tool is based on try and error to identify patterns of letters that are typical for one or the other variety. The tool was later on developed further in order to annotate messages in other national languages within the corpus, too.

5. Additional annotations

While the doctoral students were already working on their respective data, additional annotations were added to specific parts of the corpus depending on the need of the students, resulting for example in part of speech annotations for sub-corpora or specific annotations for ellipses annotated by the students themselves.
At a later stage of the project, a group of computational linguists applied neural networks (Stark et al., 2019) and encoder-decoder methods (Lusetti et al., 2018) in order to normalize the Swiss dialect into Standard German. This experiment was only carried out on a sub-part of the corpus but proved to be very successful.

6. Discussion

As has been mentioned before, several people asked about our methodology because they consider creating a collection of their own. I would thus like to mention a few points that need a second thought.
The first one is certainly the changed view on WhatsApp itself. At the time of writing, lots of discussion are going on about the privacy and the use of data by Facebook, the owner of WhatsApp. Are people still using WhatsApp for their communications or are they moving away to Signal, Telegram or another tool? The future will tell, but it seems important to realize that we performed our collection in a similar situation, when WhatsApp was bought up by Facebook and everybody was up in arms. People moved to other tools, found that they were very lonely there because none of their friends used the same tool and then the returned to WhatsApp in time for our data collection. How the current uproar will end will be seen in the future.
An important change that tool place since our data collection is the addition of multi-colored emojis. Somebody performing a new collection has to make sure that this information is not lost in their collection. Their encoding and representation might pose a challenge.
When we performed our collection, we asked our informants not to include media such as videos or picture, a function that is available in the WhatsApp backup system. In hindsight, this decision was correct because of the research questions of our doctoral students would not have profited from such contend. However, if somebody else performs a new collection, the question has to be considered again and the extra effort it takes has to be weighed against the profit that new research questions can draw from this kind of data.
As a critique of our methodology, I would like to question the order of steps that we took. Receiving the data first and the permission later on was justified in the SMS project, because the data was collected by the provider. For the WhatsApp collection, it might have been a better approach to start on the website, thus, by asking the permission to use the data first. This would have saved some time because we did receive quite some chats without permission. Having these available on our harddisks posed a small problem of data privacy for a short time but also some extra work to the team, because these chats had to be identified and removed. I would also ask the informants in more detail how their data may be used in the future, i.e. a) for the project at hand or b) for this and other scientific projects in the future or c) for all of that plus for commercial use. We often find that big companies like Samsung are interested in the Swiss German part of the data. Not having asked the according question to the informants forces us to refuse such use of the data. With such a differentiation, sub-corpora for different uses could be made available.
Looking at the demographic data that we have available for our informants, we find that everybody filled in their age and gender (the first two questions) but then many informants stopped filling in the questionnaire. In hindsight, I would ask fewer questions and spend more time evaluation the ones that are really important.
Our approach of manually annotating the main language per chat was very efficient and precise. However, it could have been avoided if the informants had been asked the

simple question about language(s) in the corpus at the moment the data was collected. Such a step would not only have saved time, the resulting language annotations would probably have been more precise.

7. References

Cougnon, L.-A. (2008). Le français de belgique dans 'l'écrit spontane´'. approche d'un corpus de 30.000 sms. *Travaux du Cercle Belge de Linguistique*, 3, 01.

Dürscheid, C. and Stark, E. (2011). sms4science: An international corpus-based texting project and the specific challenges for multilingual switzerland. In Crispin Thurlow et al., editors, *Digital Discourse: Language in the New Media*, pages 299–320. Oxford University Press, New York, London.

Lusetti, M., Ruzsics, T., Göhring, A., Samardžić, T., and Stark, E. (2018). Encoder-decoder methods for text normalization. In *Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018)*, pages 18–28, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

Stark, E. and Ueberwasser, S. (2020). The corpus "what's up, switzerland?". documentation, facts and figures. https://www.whatsup-switzerland.ch/index.php/en/corpusen.

Stark, E., Ueberwasser, S., and Ruef, B. (2015). Swiss sms corpus: www.sms4science.ch.

Stark, E., Ruzsics, T., Lusetti, M., Göhring, A., and Samardžić, T. (2019). Neural text normalization with adapted decoding and pos features. *Natural Language Engineering*, 25(5):585–605.

Ueberwasser, S. and Stark, E. (2017). What's up, switzerland? a corpus-based research project in multilingual switzerland. *Linguistik Online*, 84/5:105–126.

Language Style Accommodation in Webcare Conversations

Lieke Verheijen, Christine Liebrecht
Radboud University, Tilburg University
E-mail: lieke.verheijen@ru.nl, c.c.liebrecht@tilburguniversity.edu

Abstract

Communication Accommodation Theory (CAT; Giles, Coupland, & Coupland, 1991) postulates that language users adjust their communicative behavior to the setting, theme, or conversation partner. Interlocutors can move closer to each other by accommodating with their verbal or non-verbal language (convergence) or can create distance by magnifying disparities (divergence). The current study examined language style accommodation (also called 'alignment') in social media messages between organizations and consumers.

The interactive online environment requires adjustments to organizations' corporate language, to accommodate to the informal register of consumers on Web 2.0 (e.g., Fournier & Avery, 2011; Kaplan & Haenlein, 2010), which has resulted in the conceptualization of a conversational human voice (CHV), "an engaging and natural style of organizational communication as perceived by an organization's publics" (Kelleher, 2009, p. 177). Previous content analyses have shown that organizations indeed adopt linguistic elements of CHV in webcare messages to consumers (e.g., Huibers & Verhoeven, 2014; van Hooijdonk & Liebrecht, 2018), but the language style accommodation of both conversation partners remains uninvestigated.

In a large-scale corpus study, we examined whether consumers and organizations accommodate their language style to each other online. With the Coosto social media monitoring tool, a sample (N = 2400 messages) was created of webcare conversations on Facebook and Twitter, involving eight organizations per medium, distributed over three industries (electronics, furniture, food). The taxonomy of Liebrecht, Tsaousi, and Van Hooijdonk (2021) with seventeen subcategories that can be classified to three CHV tactics (Personalization, Informal language, and Invitational rhetoric) (van Noort, Willemsen, Kerkhof & Verhoeven, 2014) was applied to all consumer and organizational messages in webcare conversations to explore whether, how, and when consumers and organizations accommodate their language style to that of their interlocutor. Particular attention was paid to accommodation with emoji, a salient aspect of informal language. Findings and implications will be presented.

Keywords: webcare, social media, communication accommodation, alignment, conversational human voice, emoji

References

Fournier, S., & Avery, J. (2011). The uninvited brand. *Business Horizons, 54*(3), 193–207. https://doi.org/10.1016/j.bushor.2011.01.001.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *The Contexts of Accommodation: Developments in Applied Sociolinguistics* (pp. 1-68). New York, NY: Cambridge University Press.

Huibers, J., & Verhoeven, J. (2014). Webcare als online reputatiemanagement: Het gebruik van webcarestrategieën en conversational human voice in Nederland, en de effecten hiervan op de corporate reputatie. *Tijdschrift voor Communicatiewetenschap, 42*(2), 165–189.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons, 53*(1), 59–68. https://doi.org/ 10.1016/j.bushor.2009.09.003.

Kelleher, T. (2009). Conversational voice, communicated commitment, and public relations outcomes in interactive online communication. *Journal of Communication, 59*(1), 172–188. https://doi.org/10.1111/j.1460-2466.2008.01410.x.

Liebrecht, C., Tsaousi, C., & van Hooijdonk, C. (2021). Linguistic elements of conversational human voice in online brand communication: Manipulations and perceptions. *Journal of Business Research*, *132*, 124-135.

van Hooijdonk, C., & Liebrecht, C. (2018). "Wat vervelend dat de fiets niet is opgeruimd! Heb je een zaaknummer voor mij? ^EK": Conversational human voice in webcare van Nederlandse gemeenten. *Tijdschrift voor Taalbeheersing, 40*(1), 45–81. https://doi.org/10.5117/TVT2018.1.hooi.

van Noort, G., Willemsen, L., Kerkhof, P., & Verhoeven, J. (2014). Webcare as an integrative tool for customer care, reputation management, and online marketing: A literature review. In P. J. Kitchen & E. Uzunoglu (Eds.), *Integrated Communications in the Post-Modern Era* (pp. 77–99). London: Palgrave Macmillan.

Understanding Polarization: A Case Study of Black Pete in the Netherlands

Lidwien van de Wijngaert
Radboud University
E-mail: Lidwien.vandeWijngaert@ru.nl

Abstract

Polarization is generally considered as something that needs to be resolved. To reduce polarization, it is important to understand what processes cause polarization. Earlier research has proposed several hypotheses (for example, echo chambers/filter bubble and biased information processing). Using network analysis as well as quantitative and qualitative techniques, this paper investigates these hypotheses from a social media corpus perspective. Results provide support for both hypotheses. At different moments in time, different groups express their thoughts (and counter thoughts) regarding Black Pete differently. We found that the group of supporters of Black Pete is (in numbers) much more visible than the group of opponents. An explanation for large number and strong responses from supporters of Black Pete is that this group of people has something to lose, while opponents of Black Pete have something to gain. The loss that supporters of Black Pete feel seems to be about much more than just the loss of the Black Pete figure. In recent years, people have had to deal with numerous changes that affected their personal lives. People are for instance encouraged to quit smoking through higher taxes on cigarettes, they see their tradition of lighting fireworks on New Year's Eve is being criticized by environmentalists and they must accept windmills to stop climate change. Although such changes often have the moral right on their side, they also systematically affect the identity of people. The continuous flow of required changes (and consequential loss of identity) may well be the source of polarization: people put their heels in the sand to preserve their identity. So, to combat polarization, it seems important to implement changes in a way that allows people to maintain their identity. The focus of such research should be on how people can accept change while preserving their identity.

Keywords: social media, polarization, network analysis

References

Al Amin, M. T., Aggarwal, C., Yao, S., Abdelzaher, T., & Kaplan, L. (2017, May). Unveiling polarization in social networks: A matrix factorization approach. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications* (pp. 1-9). IEEE.

Araujo, T., & Neijens, P. (2020). Unobtrusive Measures for Media Research. *The International Encyclopedia of Media Psychology*, 1-7.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., ... & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216-9221.

Bruns, A. (2017). Echo chamber? What echo chamber? Reviewing the evidence. In 6th Biennial Future of Journalism Conference (FOJ17).

Bruns, A. (2019). It's not the technology, stupid: How the 'Echo Chamber'and 'Filter Bubble'metaphors have failed us. *International Association for Media and Communication Research.*

De Dreu, C. (2009). Het nut van polarisatie in politiek en samenleving: een sociaal-psychologische verkenning. *Journal of Neuroscience*, 140-152.

Fiorina, M. P., Abrams, S. J., & Pope, J. C. (2010). *Culture War? The Myth of a Polarized America*. 3rd.

Garimella, K. (2018). *Polarization on social media*. Helsinki: Aalto University publication series.

Geschke, D., Lorenz, J., & Holtz, P. (2019). The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *British Journal of Social Psychology*, 58(1), 129-149.

Holmes, J. W., & McNeal, R. S. (2016). Social Media, Participation, and Attitudes: Does Social Media Drive Polarization?. In *(R) evolutionizing Political Communication through Social Media* (pp. 1-21). IGI Global.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098–2109.

Masip, P., Suau, J., & Ruiz-Caballero, C. (2020). Incidental exposure to non-like-minded news through social media: Opposing voices in echo-chambers' news feeds. *Media and Communication*, 8(4), 53-62.

Nguyen, A., & Vu, H. T. (2019). Testing popular news discourse on the "echo chamber" effect. *First Monday*, 24(5).

Pariser, E. (2011). Beware Online filter Bubbles. Ted.

Rodenberg, J., & Wagenaar, P. (2016). Essentializing Black Pete: Competing narratives surrounding the Sinterklaas tradition in the Netherlands. *International Journal of Heritage Studies*, 22(9), 716-728.

Söderberg, B. (2017). Inside the echo chamber: A qualitative study on anti-immigration internet media, political polarization and social trust in a fragmented digital landscape.

Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review,* 34(3), 150-160.

Sunstein, C. R. (2001). *Echo chambers: Bush v. Gore, impeachment, and beyond.* Princeton, NJ: Princeton University Press.

Towards Automatic Detection of Reported Press and Media Freedom Violations in Twitter and News Articles

Tariq Yousef [1], Antje Schlaf [2], Benjamin Bock[1], Janos Borst[3]
Andreas Niekler[3], Gerhard Heyer[1]
[1]Institute for Applied Informatics (InfAI) Leipzig, Germany
[2]European Centre for Press and Media Freedom (ECPMF) Leipzig, Germany
[3]Leipzig University Leipzig, Germany
E-mail: yousef.bock.heyer@infai.org, antje.schlaf@ecpmf.eu, borst.aniekler@informatik.uni-leipzig.de

Abstract

Press freedom is under constant and increasing attack, even in Europe. Therefore, now more than ever, it is important to monitor developments and advocate for measures to protect press and media freedom. Mapping Media Freedom is a project and platform which documents threats, violations and restrictions faced by media workers across Europe and beyond. In order to support the monitoring of incidents where the freedom of press is violated, which is currently performed manually by experts, our project aims to develop a process that automatically and systematically extracts reported incidents from online text sources, namely, tweets and news articles. This paper describes our approach, which includes data gathering, creating training data for this specific topic, including solutions to overcome obstacles in this way, and evaluating different classification models, including deep learning algorithms that have produced very promising results.

Keywords: press freedom, training data creation, neural networks, text classification, social media

1. Introduction

Press freedom is under constant and increasing attack, even in Europe. Therefore, now more than ever, it's important to monitor developments and advocate to protect our press and media freedom. *Mapping Media Freedom* (MMF)[1] is a project and platform which identifies and documents threats, violations and restrictions faced by media workers across Europe and beyond. The documented incidents include physical attacks, threats of violence made online and offline, legal actions aimed at silencing critical coverage, and moves to block access to information or incidents reporting or denying access to independent and government-critical media platforms.

This project is run by the *Media Freedom Rapid Response* (MFRR)[2] [2], a rapid response mechanism against press and media freedom violations in the European Union member states and candidate countries. It provides legal support, shelter, public advocacy and information to protect journalists and media workers. The alliance is led by the European Centre for Press and Media Freedom (ECPMF) in conjunction with ARTICLE 19, the European Federation of Journalists (EFJ), Free Press Unlimited (FPU), the Institute for Applied Informatics at the University of Leipzig (InfAI), International Press Institute (IPI) and CCI/Osservatorio Balcani e Caucaso Transeuropa (OBCT). The project commenced in 2020. It is funded by the European Commission. The alerts uploaded to the MMF platform guide MFRR to directly engage with and help at-risk journalists and media workers and provide an overview of the current state and development of press and media freedom in Europe.

The incidents are published as alerts mainly by the MFRR monitoring experts, as well as an international network of local partners. However, MMF is also a crowd-sourced platform that enables anyone to upload an alert, which is then verified by the expert network before publication to guarantee the reliability of the cases and the comprehensiveness of the published details.

In order to support the labour-intensive manual monitoring of incidents, the aim of the work presented in this paper is to develop a process that automatically and systematically monitors and detects cases, namely reports about press and media freedom violations in vast amounts of online published text sources, namely news articles and tweets. Table 1 shows some examples of such tweets or news headings which are considered as reported attacks and violations and thus of interest for the experts.

The potentially relevant content that has been detected should then be presented conveniently to the monitoring experts, who then verify the incident again and publish the alert on MMF, including all the details. The publication also triggers further response actions.

The advantage of integrating automatic extraction processes is that it can use a broader range of sources and provide a faster alert mechanism. This means that more violations can be found, and more sources can be provided for each case, allowing a more realistic and reliable assessment of the press freedom situation.
of the press freedom situation.

We structure our paper according to our procedural approach: After the initial data gathering of potentially relevant content which is described in section 3, additional filters were applied to increase the proportion of potentially relevant data during the manual creation of training data described in section 4. Next, several classification models were trained (section 5) and evaluated (section 6).

[1] https://www.mappingmediafreedom.org/

[2] https://www.mfrr.eu/

Northern Ireland arrests man over journalist Lyra McKee's murder
#Germany: Demonstrators throw rocks, firecrackers at news crews in #Berlin
Poland to Propose Limits on Foreign Media
Serbian journalist detained for coronavirus reporting
Germany: #FarRight extremist group #NSU 2.0 target #reporter in new wave of #threats; Letters show extreme hatred of women
A brave British whistleblower is facing extradition to Monaco after lifting the lid on a huge bribery & corruption scandal.
Croatian President Milanovic verbally attacked HRT journalists
Arrest warrant issued on publisher/journalist @KostasVaxevanis after publishing testimonies of police torture.
A French journalist in hijab threatened with death!
Hungary's Klubradio critical of Viktor Orban set to lose license
First death threat came in April... and then extended to her family and baby son... but the journalist "won't give up my job." Solidarity with @trishdevlin

Table 1: Examples of tweets and news headings considered as reported attacks.

2. Related work

Monitoring content published in social networks to detect abuse, harassment, or freedom violations has been the subject of several research projects. (Hewitt et al. 2016, Anzovino et al. 2018, Şahi et al. 2018, Rodríguez-Sánchez et al. 2020) worked on the detection of misogynistic language and hate speech towards women on Twitter. (O'Dea et al. 2015) developed an approach to automatically detect suicide-related tweets that raise serious concern. (Bourgonje et al. 2018, Rodríguez-Sánchez et al. 2020) aimed to detect and analyse tweets that contain racism, sexism and abusive language. (Tanev et al. 2017) developed a novel unsupervised algorithm to monitor disaster impact by detecting micro-events and eyewitness reports in news and social media.

The starting point was always to define a list of keywords, hashtags, and accounts related to the topic and to collect the relevant tweets using Twitter API. Some projects collected data for a specific period of time, whereas our project collects tweets on an ongoing daily basis.

English was the language used in the majority of the projects; moreover, (Rodríguez-Sánchez et al. 2020) focused on detecting misogyny and sexism in Spanish tweets, (Şahi et al. 2018) worked on analysing and classifying tweets related to one Turkish hashtag, and (Bourgonje et al. 2018) also included a small dataset of German tweets in their model training.

However, the absence of domain-specific labelled corpora drove many projects to manually create their own training dataset according to predefined guidelines to meet their needs. Most of the manual labelling tasks involved selecting one of two binary classes (Hewitt et al. 2016, Anzovino et al. 2018, Şahi et al. 2018, Koushik et al. 2019). However, a third label was sometimes included to reflect ambiguity or doubt (Bourgonje et al. 2018, Rodríguez-Sánchez et al. 2020, O'Dea et al. 2015).

The Support Vector Machine, Naive Bayes, and Logistic Regression were the most frequently used classical classification methods with Bag of Word and TFIDF word representation. (Rodríguez-Sánchez et al. 2020, Pitsilis et al. 2018, Badjatiya et al. 2017) experimented with *Bidirectional Long Short-Term Memory* (Bi-LSTM) deep neural networks, and with *Bidirectional Encoder Representations from Transformers* (BERT) (Devlin et al. 2019). Whereas (Anzovino et al. 2018) employed a Multilayer Perceptron Neural Network.

All the approaches listed above work on historic or archived data, whereas our crawling process runs continuously and collects almost live tweets and news multiple times a day. Our project is distinguished from others by not aiming to detect direct abuses, harassment or freedom violations but reported attacks and violations. Additionally, we explicitly focus on violations and attacks targeting a specific group, namely, media actors, which includes journalists, media workers and media companies. This means that attacks or violations targeting other persons or entities shall be excluded from our project.

3. Data gathering

The first target was to gather potentially relevant data from sources that are available online to serve as a basis for the creation of training data and provide a basis for the evaluation of the trained model on unseen content, described in the following sections. Reports about violations in a specific country are expected to be found not only in news articles published in the country where the incident happened but also in news items across other countries. Furthermore, they get published in social media that are not connected to that country or language. Twitter is a social media platform that is widely used by various kinds of stakeholders, including news agencies, journalists and media workers, organisations advocating for press and media freedom, and private individuals. This allows us to detect violations reported by official accounts with high publicity as well as personal accounts, which can be especially helpful in countries where press freedom is restricted, and violations do not get reported in publicly available news media. Thus, the data gathering focused on news published online in various countries and content published on Twitter. Furthermore, we focus on content written in English to allow a wide geographical coverage.

For the data gathering, we used the free Twitter API [3] and the free version[4] of NewsAPI[5]. In order to gather potentially relevant content, we first had to collate and define relevant filters that could be used to query the APIs. This process was carried out in collaboration with the monitoring team at ECPMF, EFJ and IPI. As a result, 147

[3] developer.twitter.com

[4] The free version of the API returns the title, short description, and only 200 characters of the article content.

[5] www.newsapi.org

keywords and hashtags were defined and used to query both the Twitter API and the NewsAPI. For the Twitter API, additional 66 Twitter accounts were defined because they frequently report violations of press and media freedom based on the experience of the monitoring experts. The keywords and hashtags were defined based on three groups: (A) hashtags that are directly related to violations of press and media freedom[6], (B) keywords and hashtags related to media actors[7] and press and media freedom[8] and (C) keywords and hashtags related to attacks or violations[9]. The keyword and hashtag filter that was implemented requires either a match from a group A element or a combined match of group B and C elements. The required combination was defined in order to exclude both general content on media actors and press freedom and any other attacks that were not related to press and media freedom. As the free Twitter API does not allow partial matches, the keywords in group B and C were further extended with word forms that we assessed as valuable[10]. The crawler runs multiple times a day and collects approximately 1,000 news articles and 4,000 tweets per day.

4. Training data creation

Because of the absence of any publicly available training dataset related to our purpose, we had to create our own training dataset. Therefore, we developed an annotation tool to facilitate and accelerate the process. The annotation tool shows the tweets and news articles that have been gathered, together with buttons to label the content as relevant, not relevant, or unsure with regard to reporting about violations of press and media freedom (Figure 1).

Figure 1: The annotation tool used to collect the training dataset.

A first manual analysis of the data revealed that the classification of the content is not as straightforward as expected. Annotation guidelines were developed and shared among the human annotators to guarantee consistent annotations based on predefined rules. Borderline cases were marked as such using the button *"I can't decide"* and were discussed and agreed upon. afterwards among the annotators. This button was also used to mark a few contents written in languages different than English or a mixture of languages in order to exclude those from the future training data.

In some cases, it was difficult to decide whether the content should be considered relevant or not. This was especially the case when people posted their opinion referring to an incident. As helpful as it might be to gain knowledge via such an opinion, we do not want to train the model to return opinions, as this would add too many entries to the output. Instead, our experience showed that the incident is usually also found in other content that is reporting about the incident.

Due to the amount of data that was collected and the large number of false positives (texts that do not report about press and media freedom violations), the process would be too time-consuming for manual review by the monitoring experts. As a result, we implemented a number of processes to analyse the tweets and news articles automatically. Based on this first stage, many tweets and news items that are not relevant can be filtered out automatically before the manual annotation. First, exact text duplicates are removed. This includes duplicate news articles, retweets, authors posting the same tweet several times, and authors copying text from other authors to post as their own tweet, which occurs more often than expected in the data.

Second, we applied the keyword and hashtag filter described above to exclude items that do not contain any of the keywords and hashtags discussed in section 3. Since Twitter API returns content beyond the defined keywords and hashtags, and NewsAPI only returns the first 200 characters of the full article, applying this filter to the texts received via the APIs will increase the amount of *relevant* items in the training dataset.

[6] e.g. #journalismisnotacrime or #JournoSafe

[7] e.g. editor, journalist, reporter, photographer, camera team, blogger, whistleblower, journalism, media company

[8] e.g. #mediafreedom, #pressfreedom, #EuropeForFreeMedia, #JusticeForDaphne

[9] [9]e.g. arrested, attacked, abused, censored, bullied, defamation, harassed, insulted, kidnapped, intimidated, threatened

[10] "attack*" cannot be defined to match any word starts with "attack" but instead need to define "attack", "attacks", and "attacked" as separate keywords.

Third, a location filter was applied to filter out only news and tweets related to MFRR countries. As the author's location might be completely unrelated to the location of the incident that is being reported in the text, we employed named entity recognition using SpaCy[11] to identify locations mentioned in the text. Next, we used OpenStreetMap[12] API to translate the detected locations to latitude-longitude coordinates, which then can be mapped to countries.

Though the training data is not meant to be restricted solely to incidents in MFRR monitoring countries, this filter helped to exclude unwanted content such as large amounts of reported verbal attacks made by former United States President Donald Trump, or attacks that happened during the Black Lives Matter protests in the United States. These led to keyword matches because the texts mentioned news agencies or journalists reporting about a person under attack, but in most cases, the person being attacked was not a media actor, and thus, the content was not related to press freedom violations. As not all locations were automatically detected, the training data also includes incidents outside MFRR countries.

Fourth, after a considerable number of items was manually annotated, a pre-classification of the unannotated tweets and news was applied, using binary logistic regression trained on the already annotated data. Tweets and news items that are predicted as *not relevant* were excluded from the annotation to further increase the low proportion of potentially relevant data. The model was retrained on a daily basis to include the new annotated items. Altogether, 6,693 news articles and 9,187 tweets were manually classified. Around 26% of them are classified as *relevant*.

The inter-annotator agreement was assessed for 997 news articles and 996 tweets classified by two different human annotators; Regarding relative agreement we achieved 84.55% for news and 86.35% for tweets, and a substantial agreement regarding Cohen's kappa (Cohen 1960) with 0.62 for news and 0.63 for tweets.

5. Training

We aimed to train a binary classification model to be able to tell if a tweet or news article is reporting about press and media freedom violations or not. For this purpose, we evaluated several classic machine learning (ML) models such as SVM, Logistic Regression, Decision Tree, and Random Forest using TF-IDF representation. Additionally, we experiment with finetuning a distilled (Sanh et al. 2019) version the RoBERTa (Liu et al. 2019) model, either by directly projecting the language model's output to the classes (Transformer) or by feeding it to an additional CNN-layer (Transformer + CNN). The structure of the CNN follows the architecture in (Kim 2014), which consists of four 1D-convolutions with kernel sizes of 3,4,5 and 6, each with 100 filters. The resulting vectors are max-pooled and used for classification. In both cases the parameters of the language model are adapted during training. Distilled models have shown to be faster and smaller than their non-distilled counterpart (in the case of distilrobertabase 35% smaller), while maintaining most of its performance. We use the Hugging-face Transformers (Wolf et al. 2020) implementation for the language model.

A few preprocessing steps were carried out before the training commenced. We converted the text to lowercase and removed URLs, emojis, user mentions, and non-alphabetic characters. We excluded from the dataset the examples where the human annotators disagreed.

The size of the training dataset was 13,809 texts (5,822 news articles, 7,987 tweets) in total. The models were trained on 90% of the training dataset and tested on the remaining 10%. In order to validate the stability of the models, we trained each model on four different training/testing splits, created randomly. We guaranteed approximately the same percentage of examples from each category in the randomly selected datasets. We averaged the resulting scores among all the runs and report them in Table 2.

The models were validated on 1,007 items of unseen data labelled manually by the monitoring experts.

6. Results

Table 2 shows the scores of the models that were trained; as we can see, the deep learning models clearly outperform the classic machine learning approaches. The CNN model achieved the highest average f1-score during the training and the validation on the new unseen texts. As it has also a high recall, it was selected to be deployed and integrated into our pipeline.

	Training			Validation		
Model	Precision	Recall	F1	Precision	Recall	F1
Logistic Regression	0.5802	0.7333	0.6479	0.9482	0.6868	0.7966
Decision Tree	0.5486	0.6494	0.5947	0.8809	0.5024	0.6398
k-Nearest Neighbors	0.6372	0.6000	0.6180	0.8799	0.4421	0.5885
Linear SVM	0.6640	0.6861	0.6749	0.9627	0.5889	0.7308
Random Forest	0.8426	0.4106	0.5521	0.9932	0.2286	0.3716
Transformer	0.7768	0.8274	0.8010	0.9736	0.7659	0.8571
Transformer + CNN	0.7769	0.8324	**0.8036**	0.9559	0.8059	**0.8785**

Table 2: Training and validation results for the proposed methods.

[11] www.spacy.io

[12] www.nomimatim.org

The models were trained on news articles and tweets together. We also trained models on news and Twitter datasets separately, but this resulted in a lower F1-score between 0.06 to 0.13 when evaluated on unseen data. Thus, we decided to employ the combined model.

7. Future work

Further work aims to improve the data gathering process and the classification. Regarding the data gathering, we plan to analyse actual valuable keywords, hashtags and accounts based on the feedback provided manually on the relevance of the content in the training data as well as in the evaluation data. This can be used to complement valuable keywords and hashtags that have not yet been used as initial filters, as well as to potentially reduce filters to the most valuable ones. Moreover, we intend to continually retrain the classification model using the experts' manual feedback to enhance the classification performance and avoid topic drift.

Acknowledgment

This work is funded by the European Commission within the Media Freedom Rapid Response project and co-financed through public funding by the regional parliament of Saxony, Germany.

8. References

Anzovino, M., Fersini, E., and Rosso, P. (2018). Automatic identification and classification of misogynistic language on twitter. In Max Silberztein, et al., editors, *Natural Language Processing and Information Systems*, pp. 57–64, Cham. Springer International Publishing.

Badjatiya, P., Gupta, S., Gupta, M., and Varma, V. (2017). Deep learning for hate speech detection in tweets. In *Proceedings of the 26th International Conference on World Wide Web Companion*, WWW '17 Companion, pp. 759-760, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.

Bourgonje, P., Moreno-Schneider, J., Srivastava, A., and Rehm, G. (2018). Automatic classification of abusive language and personal attacks in various forms of online communication. In Georg Rehm et al., editors, *Language Technologies for the Challenges of the Digital Age*, pp. 180–191, Cham. Springer International Publishing.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), pp. 37–46.

Şahi, H., Kılıc¸, Y., and Sag˘lam, R. B. (2018). Automated detection of hate speech towards woman on twitter. In *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, pp. 533–536.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Hewitt, S., Tiropanis, T., and Bokhove, C. (2016). The problem of identifying misogynist language on twitter (and other online social spaces). In *Proceedings of the 8th ACM Conference on Web Science*, WebSci'16, pp. 333-335, New York, NY, USA. Association for Computing Machinery.

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1746–1751, Doha, Qatar, October. Association for Computational Linguistics.

Koushik, G., Rajeswari, K., and Muthusamy, S. K. (2019). Automated hate speech detection on twitter. In *2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, pp. 1–4.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

O'Dea, B., Wan, S., Batterham, P. J., Calear, A. L., Paris, C., and Christensen, H. (2015). Detecting suicidality on twitter. *Internet Interventions*, 2(2), pp. 183–188.

Pitsilis, G. K., Ramampiaro, H., and Langseth, H. (2018). Detecting offensive language in tweets using deep learning. *ArXiv*, abs/1801.04433.

Rodríguez-Sanchez, F., Carrillo-de Albornoz, J., and Plaza, L. (2020). Automatic classification of sexism in social networks: An empirical study on twitter data. *IEEE Access*, 8:219563–219576.

Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2019). Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *ArXiv*, abs/1910.01108.

Tanev, H., Zavarella, V., and Steinberger, J. (2017). Monitoring disaster impact: detecting micro-events and eyewitness reports in mainstream and social media. In *ISCRAM*.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38–45, Online, October. Association for Computational Linguistics.

Author Index

Keyword Index