Proceedings of the 11th Conference on CMC and Social Media Corpora for the Humanities

Céline Poudat, Mathilde Guernut

Proceedings

of the 11th Conference on computer-mediated communication and social media corpora

September, 5-6, 2024
University Côte d'Azur
France

Céline Poudat
Mathilde Guernut (eds.)

CMC 2024

11th Conference on Computer-Mediated Communication and Social Media Corpora

Proceedings of the Conference

September 5-6, 2024

https://cmc-corpora-nice.sciencesconf.org/

Proceedings of the 11th International Conference on CMC and Social Media Corpora for the Humanities

05-06 September 2024, Université Côte d'Azur, Nice, France

Editors: Céline Poudat, Mathilde Guernut

Published by: University Côte d'Azur and CORLI

Nice, 2024

Conference website: https://cmc-corpora-nice.sciencesconf.org/

Preface

Following the great success of the tenth conference held in Mannheim, Germany, in 2023, we are very pleased to present the proceedings of the eleventh edition of the International Conference on Computer-Mediated Communication (CMC) and Social Media Corpora (CMC2024). The main focus of the conference is to explore the collection, annotation, processing, and analysis of corpora from computer-mediated communication and social media.

Our general aim is to serve as the meeting place for a wide range of language-oriented investigations into CMC and social media, drawing from linguistics, philology, communication sciences, media studies and foreign language teaching and learning, with research questions stemming from corpus and computational linguistics, language technology, text technology, and machine learning.

The 11th Conference on CMC and Social Media Corpora was held at the Maison des Sciences de l'Homme et de la Société Sud-Est (MSHS) on September 5th and 6th at the University Côte d'Azur in Nice, France. This volume contains 19 accepted papers and the abstracts of the 13 posters presented at the event. Each submission was reviewed by the members of the scientific committee. The contributions were presented across three sessions with two parallel streams, along with a poster session. They cover a broad range of topics, from corpus construction to analysis, including the methods employed in that context.

The program also included two invited talks: an international keynote by Susan Herring (Indiana University, USA), who did us the great honor of attending in person, on the pros and cons of upscaling the model of Computer-Mediated Discourse Analysis she developed; and a national keynote by Marco Cappellini (University of Lyon 1) on corpora in telecollaboration and virtual exchange. This volume contains the abstracts of the invited talks. Additionally, the conference featured a community-building and metadata session, and participants were invited to attend two training workshops on Inception and Iramuteq.

We wish to thank all our colleagues who contributed to the conference and to this volume with their papers and posters. Our thanks also go to the members of the international scientific committee and to our local organizing committee, without whom the conference would not have been possible. We would also like to express our gratitude to the MSHS for kindly hosting us. We are very grateful to the University Côte d'Azur, which provided administrative and financial support: we received several grants, from the University, the Academy of Excellence 5, the Creates Graduate School and our BCL Lab. Thanks also to the national CORLI consortium, which co-organized the conference and provided both administrative and financial support.

We hope that the Nice 2024 conference will foster vibrant exchanges and contribute to strengthening the community of researchers building and using CMC and social media corpora for research in the humanities and beyond.

Nice, August 20th 2024

On behalf of the organizing committee

Céline Poudat, Cécile Angella, Marie Chandelier, Maëlle Debard, Morgane George, Mathilde Guernut, Minerva Rojas, Simona Ruggia

Committees

Organizing Committee

Céline Poudat	BCL, CORLI
Cécile Angella	CORLI
Marie Chandelier	BCL
Maëlle Debard	BCL
Morgane George	BCL
Mathilde Guernut	CORLI
Minerva Rojas	BCL
Simona Ruggia	BCL

International Steering Committee of the Conference series

Steven Coats	University of Oulu
Julien Longhi	Cergy Paris Université
Lieke Verheijen	Radboud University
Reinhild Vandekerckhove	Universiteit Antwerpen

Scientific Committee

Adrien Barbaresi	Berlin-Brandenburgische Akademie der Wissenschaften
Michael Beißwenger	UDE
Mario Cal Varela	Universidade de Santiago de Compostela
Marie Chandelier	Université Côte d'Azur
Steven Coats	University of Oulu
Louis Cotgrove	Leibniz-Institut für Deutsche Sprache
Orphée De Clercq	Ghent University
Susana Doval Suárez	Universidade de Santiago de Compostela
Annamária Fábián	University of Bayreuth
Klaus Geyer	University of Southern Denmark
Francisco Javier Fernández Polo	Universidade de Santiago de Compostela
Jennifer-Carmen Frey	European Academy of Bozen
Aivars Glaznieks	Eurac Research
Jan Gorisch	Leibniz-Institut für Deutsche Sprache
Iris Hendrickx	Radboud University
Axel Herold	Berlin-Brandenburgische Akademie der Wissenschaften
Laura Herzberg	Universität Mannheim
Mai Hodac	Université de Toulouse
Pawel Kamocki	Leibniz-Institut für Deutsche Sprache
Alexander König	CLARIN ERIC
Florian Kunneman	Vrije Universiteit Amsterdam
Marc Kupietz	Leibniz-Institut für Deutsche Sprache
Gudrun Ledegen	Université Rennes 2
Els Lefever	Ghent University
Julien Longhi	Cergy Paris Université
Paula López Rúa	Universidade de Santiago de Compostela
Harald Lüngen	Leibniz-Institut für Deutsche Sprache

Jean-Philippe Magué	ENS Lyon
Elsa María González Álvarez	Universidade de Santiago de Compostela
Maja Miličević Petrović	University of Bologna
Nelleke Oostdijk	Radboud University
Ignacio Palacios Martínez	Universidade de Santiago de Compostela
Céline Poudat	Université Côte d'Azur
Thomas Proisl	Friedrich-Alexander Universität Erlangen-Nürnberg
Ines Rehbein	Universität Mannheim
Sebastian Reimann	Ruhr-Universität Bochum
Minerva Rojas	Université Côte d'Azur
Simona Ruggia	Université Côte d'Azur
Tatjana Scheffler	Ruhr-Universität Bochum
Stefania Spina	Università per Stranieri di Perugia
Egon W. Stemle	Eurac Research
Angelika Storrer	Universität Mannheim
Caroline Tagg	The Open University
Ludovic Tanguy	Université de Toulouse
Erik Tjong Kim Sang	Netherlands eScience Center
Simone Ueberwasser	University of Zürich
Reinhild Vandekerckhove	Universiteit Antwerpen
Lieke Verheijen	Radboud University
Ciara Wigham	Université Clermont Auvergne

Table of Contents

Keynote Speakers

Talks

Posters

KEYNOTE SPEAKERS

Corpora in telecollaboration and virtual exchange

Marco Cappellini
Université Lyon1 and ICAR Laboratory
marco.cappellini@univ-lyon1.fr

Abstract

Telecollaboration is a pedagogical practice in which groups of learners in different geographical locations are connected to pursue various objectives. These objectives often include intercultural competence, soft skills such as collaboration, and linguistic skills. Since its inception, telecollaboration has been the focus of empirical studies, particularly on the online interactions among learners.

In this talk, I propose to explore corpora in telecollaboration from two perspectives. After an introduction that delimits the field of inquiry to telecollaboration, the first part will explore how researchers have collected and analyzed data throughout the history of telecollaboration. I will emphasize how researchers adapted to technological changes, moving from written asynchronous exchanges to current exchanges distributed across platforms. These platforms include asynchronous written communication, online social media, and synchronous audio-visual communication. I will also show how qualitative studies predominate in the field and point out relevant exceptions.

The second part of my talk will summarize the studies in telecollaboration that specifically deal with corpus building and analysis, especially (semi)automatic analysis. The presentation will conclude with an opening and a proposal to study multimodality in corpora of videoconference-based telecollaboration, based on my experience with the Vapvisio project...

Keywords: telecollaboration, virtual exchange, corpus, multimodality

References

Aranha, S., & Wigham, C. R. (2020). Virtual exchanges as complex research environments: facing the data management challenge. A case study of Teletandem Brasil. Journal of Virtual Exchange, 3, 13-38.

Cappellini, M., Holt, B., Bigi, B., Tellier, M., & Zielinski, C. (2023). A multimodal corpus to study videoconference interactions for techno-pedagogical competence in second language acquisition and teacher education. *Corpus*, *24*. https://doi.org/10.4000/corpus.7440

Guichon, N. (2017). Sharing a multimodal corpus to study webcam-mediated language teaching. *Language Learning & Technology*, *21*(1), 55-74.

Helm, F., & Dooly, M. (2017). Challenges in transcribing multimodal data: A case study. *Language Learning & Technology*, *21*(1), 166-185.

Large-Scale Analyses of Small-Scale Research Questions: Pros and Cons of Upscaling Computer-Mediated Discourse Analysis

Susan C. Herring
Indiana University, Bloomington USA
E-mail: herring@indiana.edu

Abstract

Advances in machine learning and generative AI mean that larger CMC corpora can be collected, normalized, analyzed, and managed more efficiently and accurately than ever before. However, in computer-mediated discourse analysis (CMDA), it is generally accepted that certain research questions are best suited for small datasets and manual analysis, due to the need for detailed, nuanced understanding and interpretation of discourse in context (Herring, 2004). Yet automated analyses of large datasets are increasingly addressing traditional CMDA topics such as identity construction, turn-taking, misunderstandings, speech acts, (im)politeness, power dynamics, and humor, and the methods for doing so are getting more sophisticated. In this talk, I will consider what CMDA gains through the latter approach, as well as what is lost if the work required to do CMDA becomes fully automated. I will argue for a mixed methodological approach, while acknowledging that AI-driven change to CMDA is probably inevitable. I will conclude by proposing that the trade-offs involved can be understood by analogy to the invention of writing, the printing press, and the internet, each of which resulted in the loss of older communication practices but represented a significant leap in how humans create, share, and interact with information.

Keywords: computer-mediated discourse analysis, large-scale analysis, manual analysis, mixed methods

References

Herring, S. C. (2004). Computer-mediated discourse analysis: An approach to researching online behavior. In S. A.Barab, R. Kling, & J. H. Gray (Eds.), *Designing for virtual communities in the service of elarning* (pp. 338-376). New York: Cambridge University Press.

TALKS

Annotating the IDA Corpus: Misogynistic and Sexist Content Across Two Transnational Incels' Communities

Selenia Anastasi
University of Genoa, Doctoral School of Digital Humanities
selenia.anastasi@edu.unige.it

Abstract

This paper presents the development of the annotation of *IDA - Incel Data Archive* (Anastasi et al., 2023), which has been collected until March 2023 from two main Incelâs fora in Italian and English. In this paper we describe a work in progress project on the annotation of textual content in both languages using a novel annotation schema. The scheme has been developed using a bottom-up inductive approach that captures distinctive elements such as multimodality, incel slang, and its misogynistic and stereotypical discourses. The result is an annotated sample of the corpus consists of a total of 3000 posts randomly extracted in both Italian and English. Finally, this paper illustrates the process of developing the annotation guideline, including an auto-ethnography description of the main challenges encountered during the annotation phase. With this work we also aim include underrepresented phenomena of online discrimination and hate speech such as body shaming, moral/slut shaming and objectification. These challenges include the conceptual definition of online misogyny and stereotypes, as well as the annotation of data collected from transnational online extremist groups.

Keywords: Comparable Corpora, Annotation, Social Media, Online Misogyny, Incel Dataset

1. Introduction

In recent years, the problem of countering the spread of online misogyny and hate speech has been approached from a variety of perspectives and research fields, from digital sociology and discourse studies to NLP and corpus linguisticis (KhosraviNik and Esposito, 2018; Fersini et al., 2018; Flynn et al., 2021). Although many monolingual resources have been developed for English, annotated resources for underrepresented languages such as Italian remain scarce. A major problem is finding data sources from social media, as privacy policies and the privatization of platforms do not allow for easy data mining. In addition, upstream content moderation within mainstream social networking sites challenges the naturalness of discursive content, the amount of representative samples of hateful speech in context, and the multifaceted complexity and flexibility of Internet language. Moreover, in recent years, a phenomenon known as networked misogyny (Banet-Weiser and Miltner, 2016) has exploded online, attracting the interest of the multidisciplinary academic community around a specific extremist subculture of the Internet known as Manosphere. The theoretical framework of networked misogyny is useful for understanding the idiosyncrasies but also thematic continuities among these groups. Recent literature has shown that the Manosphere ecosystem promotes masculinist, heteronormative, and patriarchal ideologies and practices, as well as menâs rights, anti-feminism, and incitement to the persecution of women and rape (Ging and Siapera, 2018; Sugiura, 2021; Heritage, 2023; Jane, 2014; Massanari, 2017; Nagle, 2017; O Malley et al., 2022). Such instances are connected to the Redpill set of ideas shared by users of the Manosphere. These set of ideas present an inverted narrative of the political claims of social minorities, such as Black rights movements, intersectional feminism, and other left-leaning groups, positioning cis and heterosexual men as victims of an alleged conspiracy designed to oppress them. To effectively understand and counter these web-based movements and the consequent networked misogyny, a first step is collecting data from different sources and languages, considering local peripheral groups far and beyond the Anglocentric perspective. Thus, this study aims to fill this gap by annotating a subsample of 3000 posts both in English and Italian from the larger IDA - Incel Data Archive (Anastasi et al., 2023) using a comparative bottom-up approach to the development of the categories. This research not only expands the understanding of networked misogyny from a comparative perspective, but also aims to contribute to the development of more effective and culturally contextualized responses to its harmful practices.

2. The trouble to define misogyny

Indeed, an increasing number of definitions of online misogyny have been coined in the academic literature to disambiguate and make more explicit the boundary that separates online misogyny from other forms of abuse and discriminations. In natural language processing annotation tasks, misogyny is often loosely defined as a subcategory of hate speech directed at women (Fersini et al., 2018; Fersini et al., 2022; Pamungkas et al., 2020; Guest et al., 2021; Tontodimamma et al., 2023; Richter et al., 2023; Zeinert et al., 2021). As can be seen from the summary table 1, there is general confusion in Computer Science domains about the meaning of the terms misogyny and sexism, which are often too broad or used interchangeably as synonymous. It is also worth noting that, while misogyny is legally defined as a crime, sexism *per se* is a form of prejudice that is culturally rooted and difficult to counteract with top-down methods based on lexicogrammar. On the other hands, terms commonly used in more theoretical fields are more nuanced, including the concepts of violence against women (Kilpatrick, 2004), menâs violence against women (Flynn et al., 2021), gender-based violence against women (Russo and Pirlott, 2006), sexual violence (Powell and Henry, 2017), cyber-violence against women and girls , technology-enabled violence and gendered e-bile (Jane,

2014; Jane, 2016; Massanari, 2017) â with this last scholars highlighting the role of algorithms and technological media in encouraging hateful practices.

Terms	Definitions
Misogyny	Subcategory of HS and prejudice against women or a closely related gendered group (i.e. feminism).
Misogyny as violence against gendered targets	Hateful and aggressive language based on stereotypcal gender roles, sexual orientation, gender-based hating, homophobia and transfobia
Misogyny as oppressive speech	A subtle and hidden form of Hate Speech.
Sexism	A prejudice or discrimination based on a person's gender. It is based on the belief that one sex or gender is superior to another.
Benevolent and Hostile Sexism	There exist two related but opposite orientations towards a particular gender: Hostile, which is usually harsh, angry and explicitly negative, and Benevolent, which consists in a subjectively positive view towards men or women

Table 1: A non-exhaustive collection of terminology related to misogyny largely used in the field of NLP.

On the other hand, Feminist Studies, particularly those concerned with the conceptual definition of misogyny, have challenged the mainstream definitions and dictionaries of the language in use, as they do not address the political aspects involved in the term. An exhaustive review of the theoretical literature about misogyny and sexism is beyond the scope of this paper. However, it is useful to outline some of the most influential contemporary studies. The definition of misogyny provided by Manne (2017) emphasises the active role of the social context in constructing an environment in which women are subjected to patriarchal norms of behaviour and expectations, and ultimately to various forms of hostility. According to Manne, misogyny functions to maintain a *hierarchy of power* in which women are supposed to occupy the lowest rank. In Manne's words, "misogyny is, so to speak, the police force of patriarchy, correcting women who break its laws - it functions to put women back in 'their place'" (Manne, 2017). Furthermore, her work attempts to provide an explanation for the difference between sexism and misogyny, with the former phenomenon providing the rational assumptions in the service of the latter. According to Manne, therefore, the useful approach to defining what misogyny is, should not be descriptive (dictionary approach) but rather normative. Building on the work of Manne, and a wide range of feminist scholars such as McKinnon and De Beauvoir, Richardson-Self (2018; 2021) further elaborates the concept of misogyny by suggesting that it usually can occur in two forms, which she calls *interdivisional* and *intradivisional* misogyny. The first form corresponds to hostile and coercive discourses that target women as a group. In practice, interdivisional sexist discourses can escalate into misogyny, and its conceptualization is in line with the common-sense definition of misogyny as universal hatred towards (all) women (Richardson-Self, 2018; Richardson-Self, 2021). This kind of universal discourses share several characteristics with other hate speech forms (e.g. anti-Semitism, racism, homophobia). However, Richardson-Self notes that the most common form of misogyny is *interdivisional*, which threatens only a particular type of woman: those who do not conform to the patriarchal normative order, and therefore dividing the category of women into *good women* and *bad women*. This raises the question of whether misogynistic speech should be seen as a particular type of hate speech (the so-called gendered hate speech), or whether its inherent characteristics make it a unique form of oppression. Under this perspective, Emma Jane defines online misogyny through the terms *gendered e-bile*. Through her lens, gendered e-bile suggests that, in online environments, the already restrictive social categories of *good* and *bad* women are further compressed. Indeed, even those who meet patriarchal criteria can be target of hate. More recently, finally, on the linguistic front, Deborah Cameron (2023) notes that contemporary dictionary definitions of misogyny and sexism are descriptive. That is, they attempt to capture the way words are used by members of a particular linguistic community, and from corpora. As a result, these definitions may not capture all the possible shades of meaning of a term. In her words, "Linguistically speaking, it is normal for words to be used in slightly different ways by different groups of people, for their meanings to change over time, and for their use to be affected by political differences and conflicts" (Cameron, 2023). Nevertheless, it is important to consider Feminist Studies because they offer a way to delve into how the influence of power relationships and cultural norms can affect our everyday use of language. In the next section, we propose our definition of misogyny and discuss the large and multidisciplinary body of literature that was considered in the development of the taxonomy, including the work of the scholars discussed so far.

3. Towards a working definition of online misogyny

In order to determine what could or could not be considered online misogyny, several perspectives have been considered: the existing legal literature in the European framework, relevant academic literature in sociology, philosophy, and/or linguistics, as well as the definitions provided by Italian and English descriptive commonsense-oriented dictionaries (i.e., the Oxford English Dictionary

and the Dizionario Italiano De Mauro online[1]). Regarding the European legislation, the report *Cyber violence and hate speech online against women* of the European Parliament's Department for Citizens' Rights was considered[2]. The report highlights the inconsistencies and shortcomings in the definition of many forms of violence against women. The Istanbul Convention, introduced in 2011, defines gendered violence as "a violation of human rights and a form of discrimination against women and shall mean all acts of gender-based violence that result in, or are likely to result in, physical, sexual, psychological or economic harm or suffering to women" (p.13). Otherwise, in 2018, the Council of Europeâs Additional Protocol to the Convention on Cybercrime defines sexist hate speech as "expressions which spread, incite, promote or justify hatred based on sex". While the Istanbul Convention definition emphasizes the *perlocutionary force* of these acts (i.e., their impact and possible repercussions on the target), the 2018 definition focuses instead on the *illocutionary force* of the utterance itself, stressing the responsibility of the perpetrator as well as the potential and inherent toxicity of the message. With regard to the academic literature, we considered the definitions of sexism and misogyny provided by more theoretical frameworks (as discussed in the previous section), thus considering multiple forms of harm (i.e. direct and indirect, physical and psychological) that make Internet spaces less egalitarian and safe for women and girls. The choice of combining more theoretical literature with operational definitions from the European legal framework was made in order to develop a systematic task such as corpus annotation and the development of a working definition of online misogyny:

> We define online misogyny as the heterogeneous and interconnected set of language, actions and content disseminated and mediated through digital environments whose primary aim is to oppress women (whether understood as a group or as individuals). Online misogyny is often supported and motivated by a system of both negative and positive beliefs, stereotypes and biases that justify and naturalise a hierarchy of power in which women occupy the lowest level.

This operational definition, in addition to acknowledging the multiple forms in which misogyny can occur in online environments - as observed in the literature - is consistent with the idea that it prospers within a broader socially and culturally-constructed belief system. It also shares the idea that gender stereotyping, sexism and misogyny are distinct but deeply interconnected phenomena, according to relevant feminist literature (Manne, 2017; Richardson-Self, 2018). To operationalise this definition, the Manosphere ecosystem provides an ideal case study, as it is based on a set of sexist beliefs (the so-called *Redpill*). As summarised in Table 2, during the annotation process, several factors were considered: definition of the target of the message, i.e., whether the message is directed at or refers to a gendered (female) subject or to the category of women as a social group; the pragmatic content of the expression, i.e., the presence of explicitly encoded offensive content as well as that derived from logical inference. In addition to the presence of misogyny, the annotator is also asked to check for the possible presence of stereotypes and sexism through a binary yes/no coding.

Annotation unit	Post level annotation (maximum length 281 char., calculated on average post lengths in both fora)
Subject target of the unit	A single woman or women as a group.
Pragmatic elements	Explicit vs implicit misogyny (binary annotation, mutually exclusive)
Sociolinguistics elements	Conventional vs unconventional terms (use of slang terms, binary annotation).
Multimodal elements	Presence of multimedia items in combination with texts (binary annotation)

Table 2: Main features considered during the development of the annotation task

4. Designing and Applying the Schema

The annotation task is confronted with the prevalence of newly coined words and community slang in the language, as well as the ideological framework that supports its discourse and practices. Consequently, the use of existing classifications would have resulted in the loss of the distinctive linguistic features of the dataset. Nevertheless, some of the classes introduced in our guide represent the endeavour towards alignment with other resources for hate speech (Fersini et al., 2018; Zeinert et al., 2021). Three annotators (two women and one man) experienced in Translation and Gender Studies, with backgrounds in Communication, were involved: a BA student (she/her) from Italy in Communication Studies, a PhD student (she/her) from Italy in Translation Studies and a PhD student (he/him) from Slovenia in Language Technology. All annotators are between 22 and 33 years old and are familiar with Critical Discourse Studies. One of the annotators also has been identified as neurodivergent. This aspect enriched the annotation process with considerations beyond the womenâs embodied experience of violence and discrimination, ensuring that we could also consider different and more subtle forms of ableism.

After the first pilot phase, as illustrated in Fig.1, all the annotators became familiar with the content and agreed that

[1] https://www.oed.com/?tl=true and https://dizionario.internazionale.it/

[2] The European report about Cyber violence and hate speech online against women: https://www.europarl.europa.eu/RegData/etudes/STUD/2018/604979/IPOL_STU(2018)604979_EN.pdf(Lastview:16/06/2024)

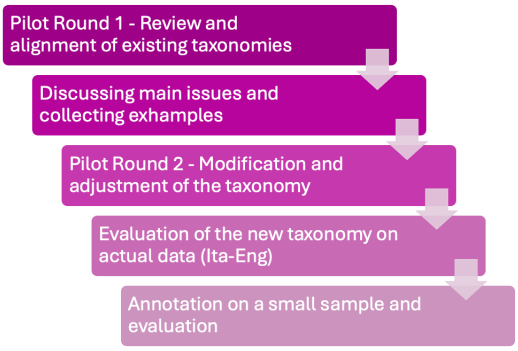


Figure 1: The iterative process of annotation.

the starting categories[3] did not capture all the characteristic features of misogyny in the incel community.
Therefore, during the second pilot round, the taxonomy was revised and extended with the following categories:
Derogation: Disparaging messages that express strongly negative connotations about women. These include explicitly offensive and demeaning terms (slurs) such as *whore*, *slut*, *cow*. Slang terms such as *femoid*, *non-person*, *roasty* are also included in this category.

IT Le NP sono NP e basta *[NP are just NP]*.

EN I try my best to avoid foids. She'll either judge me and then I'll get violent or I'll start crying.

Personal attack: A message that may fall into one or more of the previous subcategories of misogyny, but directed to a specific person. The target(s) of a personal attack must be a single person. Reference to the target may also be made through a multimedia element. Phrases that refer to multiple people, even if identified as a single person (i.e., the Maneskins, the European Parliament), cannot be considered personal attacks.

IT Onestamente vorrei vedere quale razza di uomo sceglierebbe una brutta come Susan Boyle. *[Honestly, I would like to see what kind of man would choose an ugly woman like Susan Boyle]*.

EN She really is so fucking disgusting, like I'm pretty sure all of us would rather die a virgin than fuck that thing that is a "female" [LINK TO MULTIMEDIA].

Body Shaming: Messages that express derision and denigration toward someone because of her physical appearance. Physical character is targeted because it is considered not to adhere to the aesthetic standards of the dominant culture of reference.

IT Ammesso possa chiamarsi fortuna scoparsi un'obesa pretenziosa *[Assuming it can be called luck to fuck a pretentious obese woman]*.

EN Can't even find a guy with a gf who mogs him but these landwales and chimpanzees walking around with model year guy...

Moral Shaming: Messages that blame women for certain sexual behaviours or desires that deviate from traditional or orthodox gender expectations, or that may be considered *against nature* or against religious beliefs.

IT Non é colpa nostra, le donne lo fanno apposta perché sono malvage. *[It is not our fault, it is women who do it on purpose to make us suffer because they are evil]*.

EN Why tie yourself down to a woman without a precious hymen?

Violence: It covers different types of violence, from psychological violence to explicitly expressed physical violence. Messages that promote, incite, express a willingness to commit, or show approval for acts of violence, up to and including murder, sexual assault, non-consensual penetration and harassment. Incitement to invasion of privacy is also considered an act of violence.

IT Stupida puttana spero che un giorno troverai qualcuno che ti sfondi il culo fino a farti piangere. *[You stupid bitch, I hope that one day you will find someone who will break your ass until you cry.]*.

EN It wouldn't concern me at all if I saw her being raped or mutilated in an alley.

Objectification: A central notion in Feminist Theory[4]. It includes messages associating women with inanimate objects and can be declined into: instrumentality, denial of subjectivity, interchangeability, appropriation, reduction to the body and appearance.

IT Col cazzo che vado sotto i ferri a rischiare la mia vita per un buco in mezzo alle gambe.*[I'm not going under the knife and risking my life for a hole between two legs.]*.

EN I only agree about us not letting toilets determine our worth.

Finally, to capture more characteristics of the slang, annotators are asked to answer the following questions by ticking the *yes/no* boxes:

Q1 Is there a community slang term in the post? (Examples of slang terms are Stacy, Chad, np, foid, ipergamare, etc.);

Q2 Do you have to make a logical inference to understand the sexist/misogynistic content of the post? (Refers to implicit vs explicit statements);

[3]The source categories were derived from state-of-the art literature in the tasks of linguistic annotation of misogyny, in particular the works of (Zeinert et al., 2021) and (Fersini et al., 2018)

[4]`https://plato.stanford.edu/entries/feminism-objectification/` (Last view: 15/04/2024)

Q3 Do you need to consult external sources to understand the slang term? (If the annotators need to consult resources such as the IncelWik[5] or KnowYourMeme[6] to understand the meaning of the post.

Q4 Is there a numerical rating in the post that refers to the appearance of a subject? The use of numbers and rating scales to discuss the aesthetic appearance of a subject is a practice shared by both communities (i.e. *Even 5-6/10 ethnic males will mog you out of jap/chink/Korean foids if you're a manlet sub4 white male*).

Q5 Is there any multimedia content cited or included in the piece? (Whether used to complete or emphasise a sentence or not).

These questions were included in the annotation process to ensure a quantitative assessment of the content in terms of combination between stereotypes, misogyny and conventional/unconventional speech. Q2 also allows us to understand the extent to which incel slang circulates within mainstream social networks and in contexts where the negative valence of terms is resemantised by common internet users to express other meanings, as seems to be the case with the term Chad. Our initial fundings are summarised in Table 3 below, which show the comparison of the percentage for each categories annotated in the two samples of 3000 posts in Italian and 3000 posts in English.

Categories	Italian	English
Stereotypes	$38,3\%$	$43,8\%$
Sexism and Misogyny	$73,38\%$	57%
Stereotypes using slang terms	$50,92\%$	$68,4\%$
Explicit messages	$71,80\%$	$78,9\%$
Stereotypes expressed in numbers	$14,63\%$	$7,53\%$
Unknown slang terms	$10,26\%$	$20,81\%$

Table 3: A non-exhaustive collection of terminology related to misogyny largely used in the field of NLP.

According to our manual annotation, in our balanced samples, the combination of stereotypes and misogyny is notably higher in the Italian sample (73.38%) compared to the English sample (57%). Similarly, the presence of unknown slang terms is significantly higher in the English sample (20.81%) compared to the Italian sample (10.26%). The preliminary results of the annotation highlight pressing issues of intersectionality, particularly in relation to hatred directed at women belonging to multiple marginalised groups, such as Asian and African women (particularly in the Anglophone group). In addition, the discourse reveals a significant use of ableist and ageist language. This multilayered discrimination highlights the complexity of misogynistic rhetoric within these communities and demonstrates how multiple axes of identity (ethnicity, disability, age) intersect to create unique forms of oppression and marginalisation. In the future, we plan to apply our annotation scheme to a larger sample size using an automated human-in-the-loop approach. This will improve the scalability and accuracy of our analysis, allowing us to capture a broader and more nuanced range of misogynistic discourses and related phenomena. For example, future analysis could delve deeper into the discursive modes through which intersectionality is articulated within incel and redpill cases in a comparative cross-cultural perspective. This includes analysing not only the explicit content, but also the implicit biases and structural factors that perpetuate discrimination.

[5] `https://incels.wiki/w/Main_Page` (Last view: 16/06/2024).

[6] `https://knowyourmeme.com/`(Last view: 16/06/2024).

5. Acknowledgements

I would like to thank my student Caterina Patrone for her invaluable assistance throughout the annotation phase of the dataset. This project would not have been possible without her invaluable help.

6. References

Anastasi, S., Fischer, T., Schneider, F., and Biemann, C. (2023). IDA - Incel data archive: a multimodal comparable corpus for exploring extremist dynamics in online interaction. *14–15 September 2023, University of Mannheim, Germany*, page 23.

Banet-Weiser, S. and Miltner, K. M. (2016). # masculinitysofragile: Culture, structure, and networked misogyny. *Feminist media studies*, 16(1):171–174.

Cameron, D. (2023). *Language, sexism and misogyny*. Taylor & Francis.

Fersini, E., Rosso, P., Anzovino, M., et al. (2018). Overview of the task on automatic misogyny identification at ibereval 2018. *Ibereval@ sepln*, 2150:214–228.

Fersini, E., Gasparini, F., Rizzi, G., Saibene, A., Chulvi, B., Rosso, P., Lees, A., and Sorensen, J. (2022). Semeval-2022 task 5: Multimedia automatic misogyny identification. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 533–549.

Flynn, A., Powell, A., and Sugiura, L. (2021). *The Palgrave handbook of gendered violence and technology*. Springer.

Ging, D. and Siapera, E. (2018). Special issue on online misogyny.

Guest, E., Vidgen, B., Mittos, A., Sastry, N., Tyson, G., and Margetts, H. (2021). An expert annotated dataset for the detection of online misogyny. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1336–1350.

Heritage, F. (2023). *Incels and Ideologies: Exploring How Incels Use Language to Construct Gender and Race*. Springer Nature.

Jane, E. A. (2014). âyour a ugly, whorish, slutâ understanding e-bile. *Feminist Media Studies*, 14(4):531–546.

Jane, E. A. (2016). Online misogyny and feminist digilantism. *Continuum*, 30(3):284–297.

KhosraviNik, M. and Esposito, E. (2018). Online hate, digital discourse and critique: Exploring digitally-mediated discursive practices of gender-based hostility. *Lodz papers in pragmatics*, 14(1):45–68.

Kilpatrick, D. G. (2004). What is violence against women: Defining and measuring the problem. *Journal of interpersonal violence*, 19(11):1209–1234.

Manne, K. (2017). *Down girl: The logic of misogyny*. Oxford University Press.

Massanari, A. (2017). # gamergate and the fappening: How redditâs algorithm, governance, and culture support toxic technocultures. *New media & society*, 19(3):329–346.

Nagle, A. (2017). *Kill all normies: Online culture wars from 4chan and Tumblr to Trump and the alt-right*. John Hunt Publishing.

O Malley, R. L., Holt, K., and Holt, T. J. (2022). An exploration of the involuntary celibate (incel) subculture online. *Journal of interpersonal violence*, 37(7-8):NP4981–NP5008.

Pamungkas, E. W., Basile, V., and Patti, V. (2020). Misogyny detection in twitter: a multilingual and cross-domain study. *Information processing & management*, 57(6):102360.

Powell, A. and Henry, N. (2017). *Sexual violence in a digital age*. Springer.

Richardson-Self, L. (2018). Woman-hating: On misogyny, sexism, and hate speech. *Hypatia*, 33(2):256–272.

Richardson-Self, L. (2021). *Hate speech against women online: Concepts and countermeasures*. Rowman & Littlefield.

Richter, A., Sheppard, B., Cohen, A., Smith, E., Kneese, T., Pelletier, C., Baldini, I., and Dong, Y. (2023). Subtle misogyny detection and mitigation: An expert-annotated dataset. In *Socially Responsible Language Modelling Research*.

Russo, N. F. and Pirlott, A. (2006). Gender-based violence: concepts, methods, and findings. *Annals of the new york academy of sciences*, 1087(1):178–205.

Sugiura, L. (2021). *The incel rebellion: The rise of the manosphere and the virtual war against women*. Emerald Publishing Limited.

Tontodimamma, A., Stefano, A., Stranisci, M. A., Basile, V., Elisa, I., Lara, F., et al. (2023). An experimental annotation task to investigate annotatorsâ subjectivity in a misogyny dataset. *PROCEEDINGS E REPORT*, 134:281–286.

Zeinert, P., Inie, N., and Derczynski, L. (2021). Annotating online misogyny. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3181–3197.

Beware of the Hoover : examining how users of the r/BPDlovedones subreddit use language to transform personal experiences into lay diagnostic criteria of borderline personality disorder

James Balfour
University of Glasgow
E-mail: james.balfour@glasgow.ac.uk

Abstract

This article presents a discourse analysis of a 29 million word corpus of comments from the subreddit r/BPDlovedones, a forum with 91K members, which loved ones of people diagnosed with borderline personality disorder use to share experiences. Identifying keywords against a general reference corpus of Reddit comments, the analysis identifies a set of jargonistic terms circulated on the site, which do not correspond to terms used in formal diagnostic criteria (e.g. the DSM5, the ICD). The article shows the linguistic processes by which personal experiences become reified over time as 'lay medicalised symptoms' through, for instance, nominalization (e.g. *hoovered* vs. *a hoover*). These 'lay medical symptoms' are then used as a basis for users to diagnose others, with the keyword *UBPD* (undiagnosed person with BPD) signaling this process.

Keywords: Corpus Linguistics, Critical Discourse Analysis, mental health

1. Introduction

'Seeing people with BPD as *fire* instead of *monsters* helped me immensely', writes one user on the BPDlovedones subreddit 'they are dangerous and will hurt us if we get too close.' BPD is a serious health condition which affects 2% of the UK population, meaning that about 1.3 million people are currently living with the disorder (McManus et al, 2014). To put this in perspective, this is a higher epidemiology than estimates for schizophrenia, a disorder which attracts more public interest and study. Symptoms include problems regulating emotions and chronic feelings of emptiness (APA, 2013).

Stigma around psychotic mental illnesses, and BPD in particular, is common. For instance, people with BPD are often depicted as manipulative and attention-seeking (Kling, 2014). Evidence shows that these stigmatising attitudes are internalised by medical professionals themselves, who can harbour even more negative attitudes towards people with BPD than the general population (Latalova et al, 2015). Even medical professionals will often dismiss people with BPD as manipulative, attention-seeking and ultimately 'incurable' (Kling, 2016). People with BPD are a particularly vulnerable group, even among people with mental illnesses. People diagnosed are particularly sensitive to perceived abandonment (APA, 2013) and experience internalised stigma more than people with other mental illnesses (Grambal et al, 2016). Perhaps as a result, people with BPD are a group who are significantly at risk, with more than 60% of people diagnosed attempting suicide in their lifetimes (Soderholm et al, 2020). Between 8-10% of those who attempt suicide will be successful (APA, 2013).

This article presents findings from a corpus-driven discourse analysis which addresses the following research question: 'how do friends, family and loved ones of people diagnosed with BPD use language to represent people with the disorder on the r/BPDlovedones subreddit?'. In particular, it explores subtle assumptions about BPD and mental illness and how these are encoded in repetitive linguistic patterns. The study is part of a growing body of research looking at the construction of illness in online settings (e.g. Hunt and Brookes, 2020). More specifically, the study examines a large sample of the r/BPDlovedones subreddit, which is hosted by the larger forum Reddit. According to the 'about' page on the site (which currently has 6.81k members), the r/BPDlovedones offers 'a safe space for people to discuss the challenges and abuse they have endured at the hands of someone who has Borderline Personality Disorder (BPD).' The dataset is of unique interest because r/BPDlovedones offers users anonymity, allowing topics perceived as taboo or humiliating to be shared and discussed openly. Users often use the site to share personal accounts of living with or being close to someone diagnosed (or suspected of being diagnosed) with the disorder, and to negotiate their understanding the disorder (as can be seen from the initial quotation above). Rather than study linguistic framings from the general public or people diagnosed with the disorder (Dyson and Gorvin, 2017), the present study examines linguistic framings from people who have intimate experiences with people with the disorder. This is a specific area of interest because one of the key symptoms of the disorder is difficulty forming and maintaining social relationships (APA, 2013) and positive family support has been shown to improve outcomes for adolsents with BPD (Infurna et al, 2016; Whalen et al, 2014).Thus, it is of interest to campaigners and health professionals how those relationships are conceptualised and negotiated.

2. Method

To collect the data, every comment posted to the forum between Jan 2011 and Dec 2019 was scraped from the site, resulting in a corpus that is 29 million words in size. This is one of the largest, if not the largest, specialised corpus of online forum data around a specific mental health condition collected to date. The corpus was subsequently cleaned and prepared as .txt files. Users post content under a username rather than their own name. However, since the forum contains sensitive information, and there could be some similarity between a username and a user's personal name, usernames were replaced by numerical identifiers prior to analysis. All personal names in the posts themselves were

also anonymized. To identify key lexical items, I made use of a freely available reddit corpus – the Reddit Conversation Corpus – as a reference corpus, which contains textual data from 95 random subreddits between 2016 and 2018 and is 69,428,488 words in size.

As mentioned, I identified unusually frequent lexis by comparing the wordlists from the r/BPDlovedones (target corpus) and Reddit Conversations Corpus (reference corpus). For this article, only the 100 words with the highest Log-Ratio score were examined. It was ensured that all of these keywords had a log-likelihood score above 6.63, meaning that there was only an 1% chance that the observed frequency differences were due to random variation in the dataset. Keywords, identified quantitatively, were then examined in more qualitative detail through concordancing to identify any phraseological patterns.

3. Findings

This corpus-driven approach has already revealed new insights which shows how lay accounts of diagnoses can spread in online fora. One problem with online fora centred on health-related topics is that unqualified users are able to present opinions as facts (Coulson and Knibb, 2007). At best opinions can draw on folk psychology and at worst they can promulgate inaccurate and harmful stereotypes about disorders. Indeed, in many cases on r/BPDlovedones, many users seem to have an axe to grind and the ideological motivations of different users means that language is used in strategic ways to give apparent 'objective' reported events a positive or negative spin. These representations are likely to be different from how people diagnosed with BPD tend to represent themselves.

The most striking finding is that users of the subreddit make use of a wide range of jargon which are specific to site. Many of these serve as acronyms used as referential strategies for referring to people diagnosed with the condition. These include *PWBPD* (person with BPD) (n = 29,192, LR = 139.02) and *EXBPD* (ex-person with BPD) (n = 1,726, LR = 134.94). Perhaps the most interesting of these labels is UBPD (undiagnosed BPD) (n = 3,532, LR = 135.97) which is used to refer to people in users' lives who they suspect meet the criteria for a diagnosis of BPD without having been formally diagnosed. Indeed, the way in which the "undiagnosed" meaning is packaged into the term, by not only being included as an attributive modifier (rather than a predicative assertion), but being reduced to a single letter within a larger acronym, makes this meaning more covert, and makes it difficult to distinguish references to people who have and have not been formally diagnosed. Examples include '*I just realised my fiancé has many signs of uBPD*' and '*I have two toddlers with uBPD*'. Even disregarding the illogical nature of the term (BPD *is* a diagnosis, and diagnoses cannot, strictly speaking, be undiagnosed), the term is problematic because it used almost interchangeable with the term for the actual diagnosis (PWBPD). Other notable labels include the essentialising nouns *borderlines* (plural) (n = 1,709, LR = 11.98) *borderline* (singular) (n = 7,674, LR = 7.13), which serve to equate individuals with the disorder with their condition.

Other key terminology relates to diagnostic criteria from the DSM5 such as *abandonment* (n = 4,610, LR = 10.41) (*'it originates from their core abandonment'*). However, most jargonistic terms refer to diagnostic criteria of BPD but which are specific to the language of the site and are not present in formalised diagnostic criteria, including the DSM and ICD. These include *mirroring* (n = 4,610, LR = 10.41), *discard* (n = 2,629, LR = 8.43), *discarded* (n = 2,091, LR = 7.52) and *hoover* (n = 2,572, LR = 6.78). When examining these words in context via a concordance, it becomes apparent that these are terms which are part of a larger narrative, constructed by users of the site, where people with BPD are shown to follow a cycle of idealisation and devaluation. The keyword *hoover* in the early years of the corpus functions as a verb (*' she may* try to hoover you back in') and later as a noun ('I got half sucked into a hoover'). This example showcases how time-bound experiences, occurring within specific narratives told by users of the site, gradually, over time, become reified as nouns as on the subreddit. This shows a linguistic process whereby lay experiences and presented, using language, as clinical categories, which are then used by users to make lay diagnoses.

Finally, I explore collocates around the words 'self' and 'person'. Collocates of these words include 'real' and 'true' and reveal tensions in the forum in how the 'real' identities of people with BPD are presented by users. While some users present people with BPD during the devaluation stage as the 'real' self (e.g. '*I feel as though that's their true self. Disconnected, lonely, negative and at times stunningly cruel.*'; '*the "good person inside" is a trap*) others present people with BPD during the idealisation phase as being the authentic self ('*she seemed like a genuinely good person'*) . Once again, users own experiences with people with BPD are used to legitimise more global (and thereby more stereotypical) accounts of the disorder.

4. References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, Virginia: American Psychiatric Publishing.

Dyson, H., and Gorvin, L. (2017). How Is a Label of Borderline Personality Disorder Constructed on Twitter: A Critical Discourse Analysis. *Issues in Mental Health Nursing*, *38*(10), 780-790. doi:10.1080/01612840.2017.1354105

Grambal, A., Prasko, J., Kamaradova, D., Latalova, K., Holubova, M., Marackova, M., . . . Slepecky, M. (2016). Self-stigma in borderline personality disorder - cross-sectional comparison with schizophrenia spectrum disorder, major depressive disorder, and anxiety disorders. *Neuropsychiatric Disease and Treatment, 12*, 2439-2448.

Hunt, D. & Brookes, G. (2020). *Corpus, Discourse and Mental Health* (Corpus and Discourse). London: Bloomsbury Publishing Plc.

Infurna, M. R., Brunner, R., Holz, B., Parzer,

P., Giannone, F., Reichl, C. et al. (2016). The specific role of childhood abuse, parental bonding, and family functioning in female adolescents with borderline personality disorder. *Journal of Personality Disorders*, *30*(2), 177– 92.

Kling, R. (2014). Borderline Personality Disorder, Language, and Stigma. *Ethical Human Psychology and Psychiatry, 16*(2), 114-119.

Latalova, K., Ociskova, M., Prasko, J., Sedlackova, Z., & Kamaradova, D. (2015). If You Label Me, Go with Your Therapy Somewhere! Borderline Personality Disorder and Stigma. *European Psychiatry, 30*, 1520.

McManus, S, Bebbington, P, Jenkins, R, & Brugha, T. (eds.) (2016). Mental health and wellbeing in England: Adult psychiatric morbidity survey 2014.

Söderholm, J., Socada, J., Rosenström, T., Ekelund, J., & Isometsä, E. (2020). Borderline Personality Disorder With Depression Confers Significant Risk of Suicidal Behavior in Mood Disorder Patients-A Comparative Study. *Frontiers in Psychiatry, 11*, 290.

Whalen, D. J., Scott, L. N., Jakubowski, K. P., McMakin, D. L., Hipwell, A. E., Silk, J. S.et al.(2014). Affective behavior during mother-daughter conflict and borderline personality disorder severity across adolescence. *Personality Disorders: Theory, Research, and Treatment*, **5**(1), 88– 96.

Linguistic Variations within French Wikipedia

Nelly BONHOMME
CNRS, Université Lumière Lyon 2, Laboratoire Dynamique Du Langage, France
nelly.bonhomme@univ-lyon2.fr

Abstract

Wikipedia stands as a remarkable field for linguistic study. The quantity and diversity of accessible textual data types such as articles, various talk pages, and revision histories (including precise traceability of contributors and modifications) enable the investigation of several axes of linguistic variation using quantitative analysis. In this pilot study, we examine three axes of variation: thematic variation, diachronic variation, and variation according to text type. Based on a corpus consisting of 108 French articles and their associated talk pages within Wikipedia, we observe significant differences on each of these axes for a panel of linguistic indicators, such as indicators of size, lexical diversity and distribution of parts of speech. This constitute an initial step within a larger project, providing a foundation for interesting reflections to subsequently investigate the articulation between individual and collective levels of linguistic variation within a given community.

Keywords: linguistic variation, computational linguistics, Wikipedia

1. Introduction

Wikipedia is an online encyclopedia that has a considerable amount of popularity. Besides this, it also serves as a remarkable field for linguistic study. Due to its collaborative nature and transparent policy allowing access to a complete history of contributions, this field of investigation provides an ideal vantage point for the study of the relationship between populations, linguistic variation, and change. It contributes to the understanding of the complex dynamics of linguistic variation and change. Moreover, it provides a vast amount of data, enabling both qualitative and quantitative approaches, while overcoming the "observer's paradox" highlighted by Labov (1972). This context thus offers the opportunity to revise methodological approaches and re-examine knowledge from traditional theoretical frameworks in light of contemporary societal issues. Since its inception, Wikipedia has sparked numerous reactions in the public and piqued the interest of researchers in various fields (particularly sociology and information & communication sciences). In the field of linguistics, some initial studies aim to characterize this new textual genre (Buerki, 2021; Clark et al., 2009; Emigh & Herring, 2005; Tereskiewicz, 2010). Talk pages have also drawn investigations focused on, for example, the notion of conflict (Poudat & Ho Dac, 2019). From a linguistic perspective, the potential of these data brings forward many opportunities for further exploration.

2. Theoretical Framework

Variation and change are inherent properties of language. Variationist sociolinguistics has extensively investigated these phenomena and shown that variation, far from being random and incidental, is socially structured (Chambers & Chilling, 2013; Coulmas, 1998; Gadet, 2007; Labov, 1976). These works – including the seminal article by Weinrich et al. (1968) – also acknowledge the relationship between linguistic variation and change. Indeed, the existence of different linguistic variants, conceived as different ways of expressing identical referential meaning, constitutes a prerequisite for any linguistic change (Labov, 1976; Marchello-Nizia et al., 2020). Thus, language is recognized as a complex dynamic system influenced by both internal and external factors, including population structure and social interactions. However, our understanding of these phenomena, and more specifically their interaction, remains imperfect to this day.

3. Research Topics

A preliminary exploratory study has highlighted that the textual data available within Wikipedia allows for the consideration of various dimensions of variation that coexist and are intertwined. Indeed, (i) comparing texts related to various topics enables the exploration of thematic variation; (ii) studying various versions over time enables the consideration of diachronic variation; (iii) comparing articles and talk pages allows for the consideration of both text type-related variation and intra-individual variation if we select productions from the same contributors on this axis; finally, (iv) examining productions from various contributors leads to the study of inter-individual variation. In this paper, we examine the first three axes of variation mentioned (thematic variation, diachronic variation, and variation according to text type) starting from the text (understood as an article or a talk page). With these initial investigations, our objective is to analyze whether the different axes reveal linguistic variation, and, if so, of what type, in order to then compare and conduct a cross-analysis. This constitutes a primary step within a larger project. The subsequent phase will integrate the other axes of variation mentioned above and specifically evaluate the extent to which individual and collective factors contribute to the issues of linguistic variation and change.

4. Corpus

To investigate these different axes of variation, we work with a corpus of 108 articles and their associated talk pages, all in French. The articles were selected from those rated as "featured articles" (the highest evaluation level within the

encyclopedia), diversified in terms of their topics. For this purpose, we relied on Wikipedia's system of thematic groupings, consisting of 11 major groups within the French Wikipedia: Art & Culture, Geography, History, Leisure, Medicine, Politics, Religion, Science, Society, Sport and Technology. Each article can belong to one or more of these groups. In order to retain the most prototypical articles for each topic, we first performed an initial selection to isolate featured articles affiliated with only one of the thematic groups. We then selected 12 articles per group, excluding two groups (Leisure and Medicine) that, at the time of our selection, did not contain articles exclusively associated with them. From this initial pool, the final selection of articles was quasi-random, ensuring a diverse range of creation dates.

For the associated talk pages, we take the main talk pages, featured article's talk pages, good article's talk pages, and pages for assessing article eligibility, resulting in a total of 228 talk pages. Our corpus is thus made up, based on the most recent version of each text, of approximately 1,300,000 tokens for the articles and 630,000 tokens for the talk pages. Table 1 synthetizes the distribution of tokens by thematic group and type of text (article or talk page).

THEMATIC GROUP	ARTICLES	TALK PAGES
Art & Culture	145 137	46 609
Geography	207 195	62 752
History	135 706	73 399
Politics	146 420	37 816
Religion	149 490	65 479
Science	86 255	74 279
Society	187 524	161 875
Sport	163 679	41 162
Technology	82 825	65 116
TOTAL	**1 304 231**	**628 487**

Table 1: Number of tokens according to Wikipedia's thematic group and type of text (article or talk page) in the corpus

Additionally, this corpus encompasses productions from 12,490 distinct contributors, including 264 bots, and spans over two decades (2002-2024).

5. General Methodology Elements

For each of the envisaged axes of variation, we examine the texts using a panel of quantitative linguistic indicators: number and average length of tokens, number and average length of graphical sentences (defined as beginning with a capital letter and ending with a strong punctuation mark), lexical diversity (according to a type-token ratio), and distribution of parts of speech within the texts. These indicators are automatically extracted using Python libraries such as spaCy and NLTK. They are then processed using Principal Component Analyses (PCA) and Hierarchical Ascendant Classifications (HAC) to reveal relevant axes of variation.

6. Initial Analyses and Future Developments

6.1 Thematic Variation

We first examine the articles in their most recent version (i.e., synchronously), according to their affiliation with Wikipedia's thematic groups. This analysis reveals that the thematic group is a significant explanatory factor (Wilks test, $p < 0.05$) for the variances observed within the texts on the first two dimensions of the PCA, representing 35.08% of the variability of the dataset. The indicators, among those mentioned earlier, that contribute the most to these two dimensions are ratios of proper nouns, subordinating conjunctions, adverbs, verbs, determiners, prepositions and the type-token ratio.

However, the examination of the clusters proposed by the HAC suggests considering other types of thematic groupings that could also be relevant. For example, within the first cluster, we observe a high proportion of proper nouns, which could be due to the fact that the articles focus on individuals. Then we could see a "Biography" type emerging, that would constitute a thematic typology distinct from the thematic groups proposed by Wikipedia. The second cluster, for its part, presents a significantly high ratio of numbers, and, within this cluster, there are articles belonging to the thematic group 'Sport', where it is common to find sports results and dates. In this case, there could then be convergence between the topic revealed by clustering and the Wikipedia thematic group. At this stage, these remain hypotheses that warrant further investigation to determine whether there are affinities between the thematic groups of Wikipedia and the clusters proposed by HAC, or if other thematic groupings prove more relevant for exploring thematic variation. We also aim to complement this analysis by comparing it with the structure of the contributor network to consider the potential effect of this variable.

Talk pages, on the other hand, do not show significant variation related to the Wikipedia's thematic group to which their associated article belongs. This result tends to suggest that the topics addressed within the talk pages may not always be directly related to the content of the article itself. Indeed, this is what we observe when browsing them, with more general discussions occurring, such as exchanges related to the use or not of the revised orthographic norms. Other investigations are therefore envisaged on this axis, such as a topic modeling approach to identify the main themes addressed within the talk pages and to consider whether linguistic variations are observed on this basis.

6.2. Diachronic Variation

In this section, diachronic variation is examined, initially only within the articles. Indeed, the diachrony of talk pages

requires a distinct treatment, given that the evolution of talk pages consists more of a cumulation of interventions, while articles can evolve through the addition of content but also through rewriting. Here, we focus on a *relative* diachrony within articles, which corresponds to the article's own life, i.e., its evolution from its creation to the most current version[1]. This relative diachrony aims to assess whether articles vary during their elaboration according to similar linguistic patterns or not.

For the examination of this relative diachrony, it is important to be aware that some modifications may be minor, and that articles are characterized by a non-linear evolution over time, as illustrated in Figure 1, where each black point represents a version of the article. Indeed, in the article corpus under study, we observe that they generally undergo slow phases of evolution (which are visually represented by a slight slope on the article size curve) and substantial spikes of evolution over a short period of time. These "leaps" in terms of article size are typically associated either with an external event related to the article's topic (for example, for a given personality, their death) or with an action by one or a small group of contributors to improve the article's quality and elevate it to the level of a featured article. In the case of the article exemplified in Figure 1 ("Ayrton Senna"), the observed surge in evolution is linked to this second scenario.

To consider this aspect, we select ten versions per article, based on the global size across time, dividing it into ten sections. The red squares in Figure 1 represent the selected versions. This operation allows us to have an equal number of versions for each text and also favors the selection of versions during phases of significant evolution of the article.

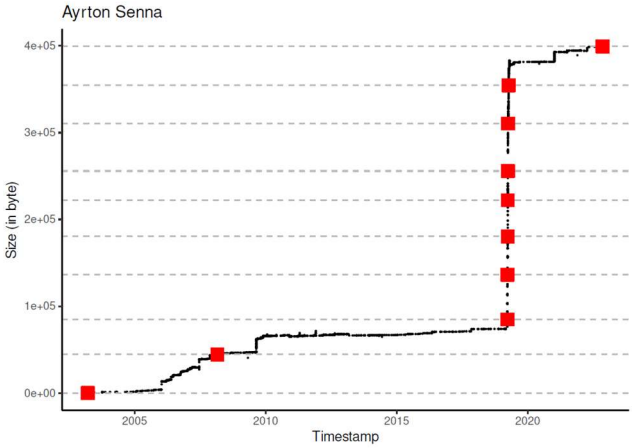


Figure 1: Evolution in size of the French article "Ayrton Senna"

The examination of the evolution of the indicators on the ten selected versions of each article reveals a significant evolution ($p < 0.05$ for all the indicators mentioned below), measured from a linear regression allowing the analysis of the significance of the evolution slope, for the majority of the indicators considered. In addition to the increase in token numbers and graphical sentences – which is intrinsic to the increase in the size of the article, we observe an increase in the average word length, lexical diversity, and parts of speech related to grammatical elements (prepositions, adverbs, coordinating conjunctions, determiners, and subordinating conjunctions). Conversely, we observe a decrease in the average length of graphical sentences and in more lexical elements such as adjectives, nouns, pronouns, and verbs. We can thus hypothesize that articles may begin with a style favoring the most meaningful lexical elements, followed by a grammatical complexification in the writing style as the article evolves. However, further analysis is needed to evaluate this hypothesis and aims to qualify these evolutions from a qualitative point of view.

6.3. Variation according to Text Types

We proceed here with a comparative examination of articles and talk pages, as well as within talk pages among different types of pages. As expected, comparing articles and talk pages as a whole sheds light on clear variation (Wilks test, $p < 0.05$ for the first two dimensions of the PCA representing 51.05% of the dataset variability), which can be related to known distinctions in terms of oral/written distinction, register, or level of formality (e.g., Gadet, 2007; Schilling, 2013). Thus, we observe the following significant differences ($p < 0.05$ for all the indicators mentioned below): articles are characterized by a larger size (number of tokens and number of graphical sentences), longer tokens and sentences, greater lexical diversity, as well as higher ratios for adjectives, prepositions, coordinating conjunctions, determiners, and nouns. Talk pages, on the other hand, present higher ratios of adverbs, auxiliaries, verbs, pronouns, and subordinating conjunctions. The results observed for talk pages also seem to be related to the use of modals, frequent among contributors, such as "I think we should better do…". Although expected, this result confirms the interest in continuing investigations by considering, on the scale of contributors, intra-individual variation that can be observed for contributors intervening jointly in articles and talk pages.

Another notable element is the observation of variation within talk pages according to the type of talk pages (Wilks test, $p < 0.05$ for the first two dimensions of the PCA representing 37.25% of the dataset variability). Further investigations are needed on this point to understand the nature of this variation.

7. Conclusion

The three axes studied highlight results confirming the existence of variation linked to various characteristics. From the thematic perspective, we observe that linguistic

[1] This *relative* diachrony is to be distinguished from an *absolute* diachrony, which corresponds to the temporal scale, independent of the life of each text. This second aspect, in turn, allows us to consider whether linguistic practices vary over time on the scale of the entire French Wikipedia. This point will not be addressed in the present work and will be subject to further investigation.

characteristics of articles seem to be influenced by the article's topic. However, it is necessary to determine whether this thematic influence aligns with the Wikipedia thematic groupings or if another type of thematic classification should be considered. Conversely, this does not seem to be the case for talk pages, suggesting that other approaches (such as topic modeling) are needed to investigate thematic variation within these texts. Regarding the diachronic axis, which we study here only in relation to the intrinsic life of the article, we observe significant evolutionary patterns in most of our indicators. Notably, there is a trend of decreasing ratios of lexical elements in favor of increasing ratios of more grammatical elements, prompting us to consider the possible existence of grammatical complexification as articles evolve. Finally, based on the type of text, we observe a clear distinction between articles and talk pages, as expected. More surprisingly, we also find significant differences between different types of talk pages, which still warrant further investigation.

The analyses presented here were initially devoted to quantitative treatment, on the scale of texts. These already provide interesting avenues for reflection that now need on the one hand to be crossed, on the other hand to be substantiated with analyses conducted on the other axes of variation considered. Especially, similar analyses are currently underway to investigate, on the scale of contributors, inter-individual, intra-individual, and diachronic variation (which will then be tackled by examining the contributors' linguistic trajectories over time). From this perspective, the originality of this work lies in the consideration of all these axes of variation within a single corpus, allowing the investigation of observed interactions. The examination of variations, both at the scale of texts (i.e., collective production) and at the scale of contributors (i.e., individual production), will also enable the exploration of interactions between individual and collective levels of linguistic variation and change, considering how a linguistic community (here, contributors to the French Wikipedia) co-constructs their linguistic conventions.

8. References

Buerki, A. (2021). What genre is Wikipedia? *Corpus Linguistics International Conference 2021 (CL2021)*, Limerick.

Chambers, J. K., Schilling, N. (Eds). (2013). *The Handbook of Language Variation and Change* (2nd edition). Malden: Wiley-Blackwell.

Clark, M. J., Ruthven, I., Holt, P. O. (2009). The evolution of genre in Wikipedia. *Journal for Language Technology and Computational Linguistics*, 24(1), pp. 1-22.

Coulmas, F. (1998). *The Handbook of Sociolinguistics*. Oxford: Blackwell.

Emigh, W., Herring, S. C. (2005). Collaborative Authoring on the Web: A Genre Analysis of Online Encyclopedias. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, Hawaii.

Gadet, F. (2007). *La variation sociale en français*. Paris: Ophrys.

Labov, W. (1972). Some Principles of Linguistic Methodology. *Language in Society*, 1(1), pp. 97-120.

Labov, W. (1976). *Sociolinguistique*. Paris: Les éditions de minuit.

Marchello-Nizia, C., Combettes, B., Prévost, S., Scheer, T. (Eds). (2020). *Grande grammaire historique du français (GGHF)*. Walter de Gruyter GmbH & Co KG.

Poudat, C., Ho-Dac, L.-M. (2019). Désaccords et conflits dans le Wikipédia francophone. In Col, G., Hanote, S. *Accord et désaccord. Travaux linguistiques du CerLiCo*, 29, pp. 155-176.

Schilling, N. (2013). Investigating Stylistic Variation. In K. Chambers, N. Schilling (Eds), *The Handbook of Language Variation and Change* (2nd edition). Malden: Wiley-Blackwell, pp. 325-349.

Tereszkiewicz, A. (2010). *Genre Analysis of Online Encyclopedias. The case of Wikipedia*. Kraków: Jagiellonian University Press.

Weinreich, U., Labov, W., Herzog, M. (1968). *Empirical foundations for a theory of language change*. Austin: University of Texas Press.

A Framework for Analysis of Speech and Chat Content in YouTube and Twitch Streams

Steven Coats
English, Faculty of Humanities, University of Oulu, Finland
E-mail: steven.coats@oulu.fi

Abstract

Online streaming platforms have become important sites of interaction and communication, but relatively little research into streaming platforms has considered the combined discourse of speech transcripts and live chat streams. In this paper we describe a pipeline approach that can integrate speech transcripts with live chat content in order to create structured documents from streams recorded on the platforms YouTube and Twitch. Built on common streaming protocols and the open-source Python library yt-dlp, the notebook comprises modular script components for data download and organization of transcripts and live chat and can additionally retrieve audio, video, and other streamed content. Additional pipeline modules can be used for automatic speech-to-text transcription of the video stream and incorporation of models for specific analytical tasks such as automatic video classification, gesture identification, or facial recognition. The paper demonstrates use of the notebook to output a time-stamped, structured combined speech/chat html file and proposes two possible analyses: consideration of chat density, and zero-shot classification of video content.

Keywords: streaming, YouTube, Twitch, multimedia, multimodal analysis, speech-to-text

1. Introduction

The increase in popularity of online streaming in the past 15 years has given rise to new, complex computer-mediated communication (CMC) environments which combine video, audio, written text, and graphical images, among other elements (Sjöblom et al. 2019). With the standardization of technical protocols for streaming and increased access to bandwidth, memory, and storage since the late 2000s, live streaming (and sharing of recordings of live streams) has become a common CMC modality on a variety of platforms which may be specialized for stream content types such as gaming and esports, talk and discussion, or other content. The most widely used streaming platforms are YouTube, which hosts a variety of streaming content, and Twitch, whose focus is primarily on gaming and esports. From the perspective of corpus-based studies of language and interaction, streams can be environments which embody multimedia and multimodal communication at multiple levels: the speech and visual content (e.g. facial expressions or gestures) of the streamer on camera, as well as, potentially, those of other persons physically or present in the same environment; the text, speech and visual content of other streams captured in the video output (in the case of Twitch, this often includes a screen showing gameplay); the text and graphical image content of the accompanying live chat, which can have hundreds or thousands of participants; and the text and graphical content of system messages such as donations or tips to the streamer, among others.

Although streaming has become a popular (and economically important) form of CMC, and a substantial research literature, particularly in computer science, has considered aspects of streaming, relatively few studies have been based on multimodal corpora which record the multiple communicative levels present in streams. High-speed chats in live streams have attracted research attention, but few studies have compared the content of live chat with the speech content of the streamer as represented in automatic speech recognition (ASR) transcripts.[1] Likewise, corpus-based comparisons of chat, transcript content, and visual or auditory content remain few.

While the modalities of this kind of interactive environment have been described in the context of CMC research, many studies have focused on disentangling the potentially complex configuration of interlocutor dynamics from a theoretical perspective, for example by describing the basic functionality of massive anonymous chat environments or analyzing aspects of online game streaming from ethnological and sociocultural perspectives. Empirical, corpus-based studies which compare the speech of the video stream, the text content of the chat, the graphicons used by participants, the automated system messages, and the content of the video stream, *inter alia*, have been relatively few, particularly from a corpus-linguistic framework. In part, this is due to the complex nature of the underlying multimodal data, which comprises a variety of video, audio, text content in different formats, which can be difficult to work with.

In this study we provide a preliminary script pipeline for capturing and combining recorded stream audio transcripts with live chat content in a timestamped tabular format that can serve as the starting point for corpus-based analysis.[2] The framework, in a notebook environment accessible via Google's Colab cloud computing environment, can also be used to collect video content. Further developments of the framework will incorporate models for various types of audiovisual analysis.

In the next section, some previous research on live streaming is presented. Section 3 describes the main elements of the pipeline and shows an excerpt from a combined speech transcript-live chat output file. In Section 4, use cases are noted: a consideration of chat density, and the potential for automated video content classification analysis. The study concludes with a brief outlook for future developments with the pipeline.

[1] See Coats (2024) for a comparison of ASR transcript content with video comments.

[2] https://t.ly/le6_e

2. Previous research

Live streams can have hundreds or thousands of participants; chat windows in live streams can therefore be fast paced and often lack interactional coherence. Herring (1999) noted that elements such as simultaneous feedback and turn adjacency can be lacking in chat environments with a large number of participants, which, however may offer possibilities for heightened interactivity and language play.

Hamilton et al. (2014) undertook an ethnographic study of Twitch gaming communities, proposing that community identity can coalesce around shared experiences in gaming streams, including in live chats. They proposed, however, that participants in massive streams with more than 1,000 viewers are focused mostly on the activities of the streamer, rather than on community interaction, which at this scale is subject to "breakdowns" due to its relative incoherence.

Ford et al. (2017) compared Twitch live chat samples from massive chats with 10,000 or more participants to samples from smaller chats, with 2,000 or fewer viewers. They found that larger chats tend to have shorter messages and more repeated content, often in the form of emotes (i.e. customized graphicon images rendered inline with chat text). Despite its seeming incoherence, "crowdspeak" can serve to consolidate in-group collective identities, for example through use of emotes or lexical items specific to a community. Harpstead et al. (2019) surveyed published research into online game streaming. They found that although many studies have been published, several desiderate remain, including "investigations that make use of broader interaction data" (p. 116).

Corpus-based studies of the language of game streaming platforms have been relatively few. Olejniczak (2015) conducted a study on a 17,500-word corpus of chat content manually collected from Twitch, finding that streams with larger numbers of participants typically have shorter chat messages. Kim et al. (2022) compiled a corpus of 15m words from Twitch stream chats to examine emotes, finding that "toxic emotes" which are used to express negative or derogatory content can be challenging to detect automatically.[3] Emotes can be used, for example, to bypass word filters or stoke racial resentments.

Recktenwald (2017) proposed a columnar transcription scheme for the analysis of the multiple communicative configurations possible in a gaming live stream: one column records timestamps, a second the speech of the streamer, the third column describes events within the game, and the fourth records chat comments. While Recktenwald used the scheme for manual transcription of speech content, this basic layout for a combined transcript is exemplified by the files generated by the automated method of the pipeline introduced in this study.

Streaming platforms have also become important economically. Streams, and recordings thereof, can be monetized by the platforms on which they are hosted. Many streamers accept donations or tips in a stream, link to paid services or e-commerce sites, or offer other kinds of paid content (Zhou et al. 2019). Johnson and Woodcock (2019) considered the economic and sociocultural implications of livestreaming of games, especially for the gaming industry itself. Yu et al. (2018) analyzed the relationship between in-stream engagement and viewer donations or gift-giving. There are few corpus-based studies of transactional events within CMC contexts, however.

In general, while a great many studies have considered aspects of streaming, the majority focus on technical considerations or larger sociocultural issues. Corpus-based studies of speech, discourse, and interaction in streams, particularly in the sense of multimodal activity, remain relatively few, in part due to the challenges inherent to wrangling the data into formats amenable to corpus analysis. Several tools exist for harvesting video and chat data from Twitch and YouTube.[4] These libraries can retrieve the JSON file of a recorded stream's live chat and render the data in various formats, but for the most part do not retrieve speech transcripts from those streams. In the following section, we describe a notebook-based pipeline for collection of chat, speech transcripts, and other data from streams.

3. Data collection pipeline

YouTube and Twitch steams are not equivalent in terms of the affordances available to streamers and viewers or the data that can be retrieved by researchers. Nevertheless, the basic structure is similar for the two platforms, and the use of common technical protocols means that the main content (video, audio, live chat) can be retrieved using functions from open-source libraries. The pipeline in this study uses yt-dlp, a Python library for the retrieval of streamed content. The steps are depicted in Figure 1.

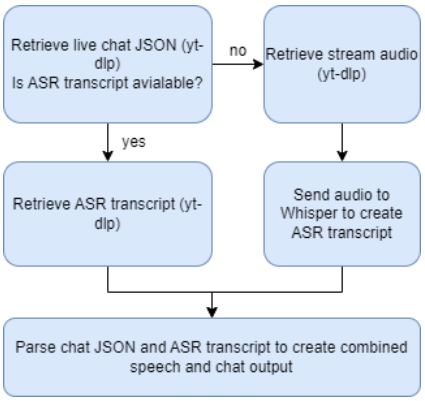


Figure 1: Flowchart of pipeline

Yt-dlp is used to retrieve the content of the live chat stream as a JSON file.[5] For streams for which an ASR transcript

[3] The authors provide examples in which an emote depicting a bucket of fried chicken, the "KFC emote", is used to make derogatory reference to racial/ethnic identity.

[4] E.g. https://github.com/lay295/TwitchDownloader, https://github.com/xenova/chat-downloader

[5] As of April 2024, yt-dlp can retrieve YouTube live chat

is available (usually most recorded videos on YouTube), the pipeline retrieves the transcript in the default VTT file format. For recorded streams for which no ASR transcript is available, the pipeline records the audio of the stream and feeds it to Whisper (Radford et al. 2022) for generation of an ASR transcript. The live chat JSON and the ASR transcript files are then parsed and the speech and chat elements rendered in a data frame in the correct order. The output is saved as an HTML file. Figure 2 depicts an excerpt from an output file, a recorded stream of the popular YouTube personality PewDiePie. The first column shows the timestamp for the utterance or chat contribution. Chatrooms can be opened prior to the start of the stream, which is why some chat entries have negative timestamps. The second column shows the transcript of the speech in the video – in this case, the speech of PewDiePie.[6] The third column shows usernames (anonymized in this screenshot), and the fourth column the chat message. The pipeline renders standard emoji (in the third row) as well as custom, non-Unicode emoji which are used in a particular channel (the emoji in the first row).

-55.000		user1	
-44.000		user2	hello rosie
-8.000		user3	wait I actually got a notification
-6.000		user4	Yesssss!!! Gonna watch AITD, live!!!
4.440	relax I'm		
8.040	early 20 minute early early gang sorry		
12.080	if anyone		
13.280	was really trying to time this it's a		
16.760	hard with a		
18.000		user5	Yey!!!!

Figure 2: Excerpt from output file

4. Use cases

An aligned transcript containing speech and chat messages can be used as the basis for investigations of the properties of multimodal CMC, including analyses of grammatical phenomena, lexis and discourse, and emoji and emotes. In addition, information from the aligned transcript can be utilized in the context of analytical steps undertaken on the underlying video data.

4.1 Chat density

One possible approach is to analyze chat density in streams and correlate stream content and/or speech messages with periods of high or low chat density. Figure 3, again from the PewDiePie stream cUUuRK3Rm4k, shows density of streamer speech and live chat messages. In this figure, the blue line represents the density of chat messages per minute, and the orange line the number of utterances per minute by the streamer. As can be seen, there is little communication in the chat stream prior to the streamer starting the stream (at 0 minutes). Thereafter, densities remain fairly constant: speech at approximately 20-25 utterances per minute, and chat between 25 and 50 messages per minute. The exception is from minute 105 until the end of the stream. Here, examining the aligned transcript provides clues: The large increase in number of chat messages is prompted by the streamer interacting directly with his audience by saying "I totally forgot there were so many people watching, hey it's good to see you guys", prompting a large number of messages and responses in the chat; this is followed by several more comments by the streamer addressing the audience directly.

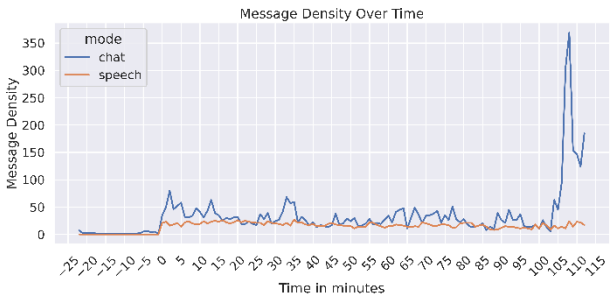


Figure 3: Chat and speech density for stream cUUuRK3Rm4k

4.2. Automated analysis of video streams

The notebook format of the pipeline may be suitable for combining text-based analysis of speech and chat messages with automated methods for the analysis of multimodal visual-textual content, for example by importing models from Huggingface. Stream segments with high levels of chat activity can be analyzed with models for zero-shot classification of visual content (Ni et al. 2022). Using individual speech turns and/or chat messages as the input texts in a classification task may provide insight into the dynamics of the underlying multimodal communication by clarifying who is commenting on the video and who is commenting on (for example) other chat or speech content. The incorporation of automated video analysis components into the Colab environment for the framework is planned.

5. Summary and outlook

CMC in the past 30 years has undergone a shift from primarily text-based modalities such as message boards or chatrooms towards multimedia environments in which video, audio, text, and images are all shared in real-time in streams. These complex environments require new methods for the organization and curation of data in corpus formats that are suitable for a variety of analytical approaches. The stream pipeline described in this paper presents a notebook environment which retrieves streamed data and combines speech transcripts with live chat messages, allowing the analysis of discourse and graphical content which prompt high rates of chat density. In addition,

streams by default, but to retrieve live chat streams from Twitch video on demand, a patch must be installed (https://github.com/yt-dlp/yt-dlp/pull/1551).

[6] https://www.youtube.com/watch?v=cUUuRK3Rm4k

the environment can integrate tools for ASR and video analysis which allow multimodal comparisons to be undertaken.

Future developments with the pipeline will have three main focuses: First, to consider streams not only from YouTube and Twitch, but from other platforms, covering different kinds of content. Second, to incorporate versatile video analysis models which can provide automatic descriptions of video content such as in-game developments. Third, to incorporate video recognition models that can account for facial expressions and gestures (Parian-Scherb et al. 2022). Models for automatic video content recognition can then be used to generate outputs which can be analyzed in the context of a stream's speech, chat, or system message content.

Advances in AI models for the analysis of video data are ongoing, and in the immediate future, automated annotation of streaming data will undoubtedly be feasible. A notebook-based data collection and analytical environment such as the one presented in this study, which allows time-stamped transcripts of speech and chat content to be combined, will provide a foundation for further developments in the corpus-based analysis and understanding of multimodal online interaction.

6. References

Coats, S. (2024). Commenting on local politics: An analysis of YouTube video comments for local government videos. *Research in Corpus Linguistics*.

Danesi, M. (2017). *The semiotics of emoji: The rise of visual language in the age of the internet*. Bloomsbury.

Ford, C., Gardner, D., Horgan, L. E., Liu, C., Tsaasan, A. M., Nardi, B., & Rickman, J. (2017). Chat speed OP PogChamp: Practices of coherence in massive Twitch chat. In *CHI EA '17: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems May 2017* (pp. 858–871). https://doi.org/10.1145/3027063.3052765

Hamilton, W. A., Garretson, O., & Kerne, A. (2014). Streaming on Twitch: fostering participatory communities of play within live mixed media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1315–1324).

Harpstead, E., Rios, J. S., Seering, J., & Hammer, J. (2019). Toward a Twitch research toolkit: A systematic review of approaches to research on game Streaming. In *Proceedings of the Annual Symposium on Computer-human Interaction in Play* (pp. 111–119).

Herring, S. (1999). Interactional coherence in CMC. *Journal of Computer-Mediated-Communication*, 4(4). https://doi.org/10.1111/j.1083-6101.1999.tb00106.x

Johnson, M. R., & Woodcock, J. (2019). The impacts of live streaming and Twitch.tv on the video game industry. *Media, Culture & Society*, 41(5), 670–688. https://doi.org/10.1177/0163443718818363

Kim, J., Wohn, D. Y., & Cha, M. (2022). Understanding and identifying the use of emotes in toxic chat on Twitch. *Online Social Networks and Media*, 27. https://doi.org/10.1016/j.osnem.2021.100180

Konrad, A., Herring, S. C., & Choi, D. (2020). Sticker and emoji use in Facebook Messenger: Implications for graphicon change. *Journal of Computer-Mediated Communication*, 25(3), 217–235. https://doi.org/10.1093/jcmc/zmaa003

Ni, B., Peng, H., Chen, M., Zhang, S., Meng, G., Fu, J., Xiang, S., & Ling, H. (2022). Expanding language-image pretrained models for general video recognition. *arXiv*, cs.CV, 2208.02816. https://doi.org/10.48550/arXiv.2208.02816

Olejniczak, J. (2015). A linguistic study of language variety used on twitch.tv: Descriptive and corpus-based approaches. In: *Proceedings of RCIC'15: Redefining Community in Intercultural Context, Brasov, 21–23 May 2015* (pp. 329–334).

Parian-Scherb, M., Uhrig, P., Rossetto, L., Dupont, S., & Schuldt, H. (2023). Gesture retrieval and its application to the study of multimodal communication. *International Journal on Digital Libraries*. https://doi.org/10.1007/s00799-023-00367-0

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust speech recognition via large-scale weak supervision. *arXiv*:2212.04356 [eess.AS]. https://doi.org/10.48550/arXiv.2212.04356

Recktenwald, D. (2017). Toward a transcription and analysis of live streaming on Twitch. *Journal of Pragmatics*, 115, 68–81.

Riddick, S., & Shivener, R. (2022). Affective spamming on Twitch: Rhetorics of an emote-only audience in a presidential inauguration livestream. *Computers and Composition*, 64, 102711.

Siever, C. M. (2019). 'Iconographetic Communication' in digital media: Emoji in WhatsApp, Twitter, Instagram, Facebook—From a linguistic perspective. In E. Giannoulis & L. R. A. Wilde, (Eds.), *Emoticons, kaomoji, and emoji: The transformation of communication in the digital age* (pp. 127–147). Routledge. https://doi.org/10.4324/9780429491757

Sjöblom, M., Törhönen, M., Hamari, J., & Macey, J. (2019). The ingredients of Twitch streaming: Affordances of game streams. *Computers in Human Behavior*, 92, 20–28.

Spina, S. (2019). Role of emoticons as structural markers in Twitter interactions. *Discourse Processes*, 56(4), 345–362. https://doi.org/10.1080/0163853X.2018.1510654

Yu, E., Jung, C., Kim, H., & Jung, J. (2018). Impact of viewer engagement on gift-giving in live video streaming. *Telematics and Informatics*, 35(5), 1450–1460.

Zhou, J., Zhou, J., Ding, Y., & Wang, H. (2019). The magic of danmaku: A social interaction perspective of gift sending on live streaming platforms. *Electronic Commerce Research and Applications*, 34, 100815.

The analysis of 'inclusion' and 'accessibility' in Computer-Mediated-Communication for an inclusive transformation in digital societies[1]

Annamária Fábián, Igor Trost, Kevin Altmann, Mara Schwind
University of Bayreuth/University of Passau/bidt – Bavarian Research Institute for Digital Transformation/University of Passau
E-mail: annamariafabian@yahoo.de, igor.trost@uni-passau.de, kevin.altmann@bidt.digital, mara.schwind@uni-passau.de

Abstract

Diversity and inclusion are crucial elements in building equitable and thriving digital societies. While digitalization provides a number of opportunities to diversity and diverse individuals via digital visibility, it is also connected with discrimination, digital divide and exclusion. Digital transformation leads to a transformation of every field of life – e.g. society, politics, law, other professional sectors of life – and with this, there is also a transformation of communication and communicative habitus. Research on computer-mediated-communication enables the monitoring of accessibility, inclusion as well as the digital divide by examining digital corpora of diverse groups and their comments on accessibility and inclusion in digital communication. This paper examines therefore disability-related diversity and inclusion of disabled communities and their communication through a digital linguistic (data-driven and data-based) approach, which will be used for empirical observations on digital societies concerning accessibility, discrimination, and inclusion but also for recommendations for an inclusive digital policy-making.

Keywords: #Accessibility, #DigitalDivide, #Inclusion, #DigitalTransformation, #DigitalLinguistics, #LinguisticsofInclusion

1. Introduction

As technology continues to play a central role in various aspects of digital societies, it is essential to ensure that the benefits and opportunities digitalization offers are accessible to and shared by everyone, regardless of their background, identity, or abilities. While digital transformation leads to rising public awareness for diversity, empowerment & inclusion, digitalization also brings risks of inequity, discrimination, and exclusion such as the digital divide to diverse and often marginalized individuals and collectives. Simultaneously, the digital participation of diverse groups via computer-mediated-communication enables the identification of barriers to be overcome for a more inclusive digital transformation. There are 7,8 million people with disability in Germany (9,4 % of the German society)[2] and 1.3 billion worldwide (16 % of the world population)[3]. In terms of the UN Convention on the Rights of Persons with Disabilities (UN CRPD) on the one hand, and the high number of individuals with disability with a rising tendency due to demographic change on the other hand, it is necessary to look into human-centered-data on inclusion and accessibility in digital societies and to set the course for an inclusive digital transformation. As a result of the democratization of digital media, people with disability use social media to raise their voices for inclusion and accessibility by identifying barriers in computer-mediated-communication as well as communicating factors essential to digital accessibility. While people with disability aim at shaping the digital transformation towards inclusion, academic research mainly focuses on gender diversity as well as participation in digital societies in general barely reaching other diversity dimensions beyond gender. Hence, disability-related diversity and other diversity dimensions in communication on social media still remains underrepresented. Consequently, there is an academic & social emergency of research on theories and methodology for exploring digital accessibility and participation of humans with different diversity dimensions. This paper is therefore dedicated to the digital participation of people with disability and their activism for digital accessibility via computer-mediated-communication, both as a research desideratum. Our interdisciplinary research is based on an interdisciplinary research design integrating theory and methods of applied communication science, sociology, and law. In this study, we first set the theoretical framework by providing definitions of `inclusion` from sociology and then analyze the use of `inclusion` in computer-mediated-communication between 2009-2023 in a digital corpus from Twitter (now X) compared to an analog corpus. After gaining a first impression of the relevance and progress of 'inclusion' in society, we focus on the participation of people with disability on social media and examine their communication on Twitter concerning digital accessibility in the time span of four months, which allows us first valuable insights into the perspectives of disabled individuals on this relevant topic. Both, applied communication research as well as law science often use an empirical analysis of key words related to emergent social issues. The benefit of this study is twofold as the key-word-analysis of 'inclusion' and 'accessibility' enables an overview of the progress of these words in society but is used among others for gaining knowledge of the social relevance of access for inclusive digital policy-making. According to

[1] This paper was written as part of the project "Communication of Disability-Related-Diversity on Social Media for Inclusion". The project is affiliated with the bidt – Bavarian Research Institute for Digital Transformation at the Bavarian Academy of Sciences and Humanities (Munich/Germany) and funded by the Bavarian State Ministry of Science and Art.

[2] https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Behinderte-Menschen/_inhalt.html

[3] https://www.who.int/news-room/fact-sheets/detail/disability-and-health

the UN CRPD, inclusion must be provided to every individual. A further unique-selling-point of this study lies not only in the interdisciplinarity between corpus linguistics, communication science, law, and social sciences (sociology) but also in the exploration of computer-mediated-communication concerning disability-related diversity and inclusion including the communication of people with disabilities and towards them, a prospective but emergent field in research on CMC. In addition, this research paper is at the intersection of CMC and Human-Centered-Data-Science. Whereas a wealth of significant studies (e.g. Herrera, 2022 and Zelena, 2020) addresses disability in digital societies, the communication on disability diversity as well as the computer-mediated-communication of people with disability is still underrepresented in corpus linguistics. While Herrera (2022) argues that social media analytics tools need to be designed to support inclusive public services for all, including those with disabilities, Sinclair (2010) emphasizes the importance of paying attention to social barriers that inhibit inclusion, rather than simply technological barriers. In addition, Zelena (2020) examines the digital divide and explores how new media platforms become the platform of communal loss for users of different ages, genders, social statuses, and diverse internet usage habits and socialization. This study, however, analyses inclusion and the disability digital divide on social media in a social and legal context and shows how people with disability and their digital communities drive inclusion in digital societies by using particular hashtags and computer-mediated-communication. First, we set the sociological framework with scientific considerations on inclusion, then we carry out a frequency analysis of 'inclusion' and 'accessibility' in a corpus from Twitter (X) in the time span of 14 years. After gaining insights into the social relevance of these definitions in progress based on a corpus-linguistic examination, we turn our attention to 'accessibility' and carry on a small study with appr. 208,829 tokens in the time span of 4 months for more details on community engagement for digital accessibility.

The author's vision behind this study is to turn discrimination or exclusion into participation & inclusion through data mining & qualitative research for the identification of factors essential to digital accessibility and a more inclusive digital transformation.

2. Inclusion in digital societies – an approach from Sociology for a theoretical framework

First, our interdisciplinary research group sets a multi-layered definition of inclusion in digital societies, which is the theoretical frame for further analysis in this paper. Inclusion as "being part of" corresponds to our common, everyday understanding of what we mean by inclusion, namely participation in societal activities, groups, or institutions. However, even in this everyday understanding, the note that inclusion can likewise differentiate you from something else or, in extreme cases, partially or completely exclude you, marks a central starting point to look at inclusion in (digital) societies.

In sociological terms, inclusion and exclusion of/in societies are therefore not necessarily opposites or cancel each other out. Rather, they mark a tense relationship with regard to (non-)belonging and participation in social processes (cf. Husemann 2017, 62). This potential approach to inclusion/exclusion "argues from an intersectional perspective rather from a social science standpoint, understands inclusion as a necessary counterpart to exclusion and analyzes questions of social inequality" (Budde & Hummrich 2015, 166). As Husemann points out, inclusion and exclusion can be understood from various sociological perspectives - albeit only cursorily at this point[4] - as conditions of participation in social action (Weber 2005), as part of the functional differentiation of society into social subsystems such as politics, the economy, education, etc. and the participation or membership in those areas (Stichweh 2009, Marshall 1964), in the context of shaping social solidarity (Durkheim 1933) or institutional disciplining (Goffman 1961) (cf. Husemann 2017, 64ff).

Comprehensively, however, it can be stated that inclusion always marks a demarcation between those included, those excluded, and actors on the peripheries of social spheres of participation. In this way, asymmetries and demarcations, sometimes intended and functional for various parties, are mutually generated as forms of social order. To summarize, this *first reading* of inclusion from a sociological-analytical perspective focuses on the interdependencies between inclusion and exclusion – on the occasion of this paper 'digital divide' –, particularly concerning the conditions for social participation and the establishment of controlling institutions for the **constitution of (non-)belonging in modern differentiated societies**.

In contrast to a sociological-analytical perspective, a *second reading* of inclusion can be identified as a demand for full and equal participation in society and the right to **self-determination** (cf. Bittlingmayer & Sahrai 2017, 686), particularly by those who are at risk of exclusion. This second reading of inclusion can be described in contrast to the first reading as **social inclusion in a normative demand** (for more on the distinction between the two readings, see Husemann 2017). Concerning social inclusion, reference is made to a *narrow* and *broad* understanding of the term. Therefore, inclusion is mostly used to refer to the participation of people with disabilities as a *narrow understanding* of the term, which aims to comply with the UN Convention on the Rights of Persons with Disabilities (UN CRPD) (cf. Bittlingmayer & Sahrai 2017, 686; cf. Budde & Hummrich 2015, 166). In this respect, the UN CRPD is primarily used as a frame of reference to analyze the concepts of social inclusion/exclusion.[5] The UN CRPD's

[4] For a further overview of sociological theories in the context of inclusion, see Husemann 2017.

[5] For example, the UN CRPD is being examined as an instrument for an anti-discrimination strategy, whereby the UN CRPD tightens the understanding of non-discrimination, but could also lead to the marginalization and invisibilization of the experience of disability through a changed normalization understanding of disability (as part of human diversity) (Bittlingmayer/Sahrai

definition of inclusion (in the sense of the narrow understanding) is as follows:

> „The Convention on the Rights of Persons with Disabilities is no longer about the integration of the 'excluded', but about enabling all people to participate fully in all activities from the outset. It is not the a priori negative understanding of disability that should be the norm, but a common life for all people with and without disabilities.“ (http://www.behinderten-rechtskonvention.info/inklusion-3693/).

Even if an introduction to a sociological perspective on inclusion could only be given here in a short overview, it was emphasized that the idea of inclusion as a „fuzzy concept“ (Artiles & Dyson 2005, 43) entails ambivalences that need to be reflected in interdisciplinary research from the point of view of Computer-Mediated-Communication, Law, and Sociology.

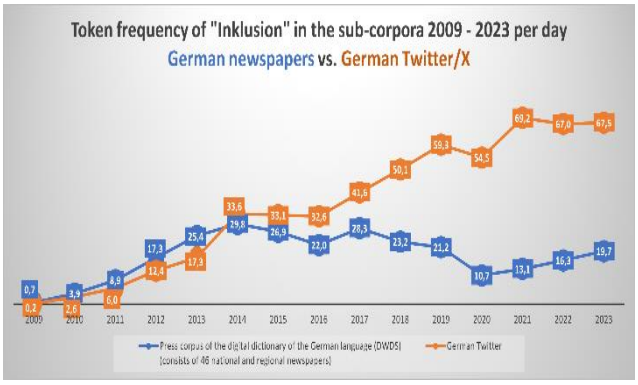
From a sociological perspective, inclusion and exclusion are not just separate terms with a normative charge, but intertwined and represent constitutive functions of modern societies for the production of social order, e.g. the self-definition and demarcation of (collective) identities or the coordination of institutional/organizational activities (cf. Scherr 2017, 47), which fundamentally affects all members of society.

Based on this sociological view, however, the perspective of marginalized actors with experiences of social exclusion should not be de-emphasized by means of the challenges of inclusion and exclusion faced by digital societies as a whole. Rather, a sociological framing of inclusion/exclusion can sharpen the analytical focus on reflecting normative demands for social inclusion, the articulation of social inequality by those affected, and the empirical observation of exclusion as ubiquitous negotiations of boundaries and „being part“ of society. These negotiations arise particularly where actors on the verge of marginalization articulate social participation as deficient and therefore make it visible through digital community engagement and computer-mediated communication. In our interdisciplinary article, we look at the articulation of the needs for social inclusion through inclusive digital communication and the exclusion experienced and called out by people with disabilities on and via social media. Regarding this, we observe the usage of social media to address digital exclusion not only as a representation of a deficient inclusion landscape in digital societies. Rather, we understand the use of social media as a strategic means of self-empowerment for those affected to position themselves as an instance for countering exclusionary communication. From there, negotiations about the limits of social participation are conducted vis-à-vis institutional representatives on social media. Against the background of a sociological perspective on inclusion/exclusion and current digital divide research, we reflect on the venues that social media offer to address and question the (non-)inclusive communication on the Internet as boundary-drawing processes between a peer community of people with disabilities and an environment perceived as ableist (cf. Haage 2020, 291).

3. ‘Inclusion` and ‘accessibility’ in the computer-mediated-communication of digital societies

As pointed out in the chapter on the considerations of ‘inclusion’ from the point of view of sociology, the empirical analysis of the CMC on ‘inclusion’ and ‘accessibility’ can be useful for observations concerning demands for social inclusion and the articulation of social inequality in societies by those affected. In order to monitor the relevance of ‘inclusion’ in German society, a corpus linguistic framework consisting of corpus-driven quantitative methods was set up. This part of the study focuses therefore first on the progress of the word ‘inclusion’ in digital big corpora from social media (Twitter/X) and digital press between 2009-2023 in comparison, which is demonstrated below in figure 1:

Figure 1: Token frequency of ‘Inklusion’



Whereas the digital press corpus was generated from DWDS (Digital Dictionary of the German Language‘, published by the Berlin-Brandenburg Academy of Sciences and Humanities), the comparative **representative digital social media corpus was collected from Twitter/X from 2009 – April, 18th 2023** (22:35:02 UTC) and it consists of **5,663,504 tokens.**[6]

The monitoring of the frequency of ‘inclusion’ in both corpora indicates the rising significance of ‘inclusion’ in digital press and digital society simultaneously, but the table also demonstrates a distribution of public awareness for inclusion from the digital press to social media. This finding can be attributed to the fact that people with disabilities recognize the potential of social media (in this case Twitter/X) for activism and actively use #Inklusion (‘inclusion’) for community organization and disability visibility. While the frequency of ‘inclusion’ had constantly been increasing between 2009 and 2021, the interest in the digital press on ‘inclusion’ remained moderate – even decreasing until 2020, the first year of the pandemic. The highest frequency on social media was achieved in 2021, which was the most emergent year of the global pandemic COVID-19 putting the lives of many individuals with disability and/or underlying health conditions at a high risk. According to Fábián

2017).

[6] Acknowledgement: Our research group „Diversity and Inclusion in Digital Societies“ at the Bavarian Research Institute of Digital Transformation (bidt) at the Bavarian Academy of Science (BAdW) would like to thank Prof. Dr. Jürgen Pfeffer (Professor for Computational Social Science) at the Technical University of Munich for collecting the corpus for us.

& Trost (2024), this health-threatening circumstance led to a high level of disability activism and community engagement via computer-mediated-communication on Twitter (X) enabling digital interactions with politicians and media for anti-discriminatory and inclusive policy-making during the pandemic. Consequently, this activism likely led to increased reporting on inclusion also in the digital corpus indicated by an increasing frequency of the use of the word 'inclusion' in the digital press corpus from DWDS. While the use of 'inclusion' on Twitter/X has decreased since 2021, it obviously has been gaining relevance in digital press, but the activism of people with disability on social media still remains significant as the CMC-focused-content-analysis consisting of Tweets along the hashtags 'Inklusion' ('inclusion') and 'Behinderung' ('disability') reveals, according to Fábián & Trost (2024), significant factors for an inclusive transformation in on- and offline life communicated directly by people with disability. The ethical and participatory approach of this study has arisen of the fact that our research is not restricted to the views of mainly scientists on inclusion, but includes disability-centered data and perspectives of people with disability on inclusion. As the corpus analysis demonstrates the rising significance of 'inclusion' in digital press corpora & corpora from social media, we carry out a CMC-analysis of 'Barrierefreiheit' ('accessibility') as well as 'digitale Barrierefreiheit' ('digital accessibility'). Afterward, we focus on the disability digital divide, which people with disability are facing and therefore countering on social media.

4. Disability activism via computer-mediated-communication for digital accessibility in pluralistic digital societies

Our corpus examination reveals that the word 'Inklusion' ('inclusion') often builds a collocation with 'Barrierefreiheit' ('accessibility') and also 'digitale Barrierefreiheit' ('digital accessibility)'. Our further corpus study is therefore based on this key finding. First, we show the increasing frequency of both words in our corpus from Twitter/X in the time span of 14 years – between 2009-2023 – underlying to the examination of 'Inklusion' ('inclusion') in the last chapter and then accomplish a small quantitative and qualitative study on 'digitale Barrierefreiheit' ('digital accessibility') for monitoring the digital disability divide and gaining further insights in this in a Twitter/X corpus of 4 months between January, 1st 2023 until April, 18th 2023 comprising 7,963 tweets in German with **208,829 tokens.** The first table demonstrates that 'barrierefrei' has gained more awareness since 2009 on behalf of the discourse on 'Behinderung' ('disability') and 'Inklusion' ('inclusion'). A relatively high significance with at least 5000 tokens per million can be observed in 2021 in the corpus, which is – as the high frequency of 'Inklusion' in the same corpus in general too – associated with COVID-19 and the emergency of accessible information in particular in times of global health threatening crisis. The lexical analysis of 'accessibility' and 'accessible' (table 1) shows the shift of social transformation towards accessibility via digital transformation by an increasing use of the word 'Barrierefreiheit' ('accessibility') "came along" since 2009 in the digital discourse on disability ('Behinderung') & inclusion ('Inklusion'):

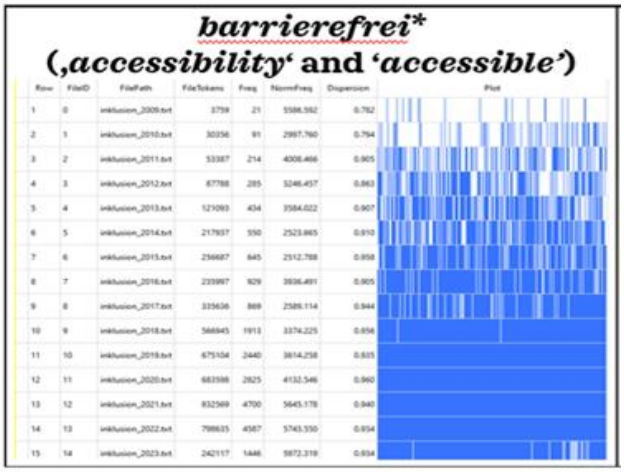


Table 1: 'barrierefrei*'

While 'barrierefrei' has been established since 2009, the second table shows that 'digitale Barrierefreiheit' ('digital accessibility') appeared in the Twitter discourse of 14 years the first time in 2013 with the broader use of Twitter/X by people tweeting in German along the hashtag 'Inklusion' and 'Behinderung'. The spreading use of social media however is associated with barriers faced by people with disability, which they seek to overcome for an inclusive digital transformation, which explains the increasing use of 'digitale Barrierefreiheit' in the corpus (table 2):

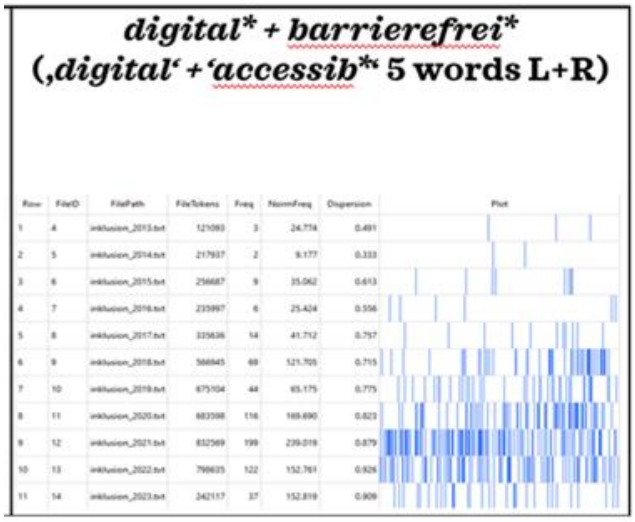


Table 2: 'digital*+barrierefrei*'

As already demonstrated in the analysis of 'accessibility', 'digital accessibility' ('digitale Barrierefreiheit') has attracted more attention than ever before in the corpus since 2020 - 150 to 240 tokens per million – attributed to the pandemic following the need for accessible digital information with risks and warning to the whole population, and of course, with even higher relevance to people with disability

and underlying conditions. Our decision to carry out a more detailed empirical analysis of the participants tweeting on '(digital) accessibility' and their community engagement against the digital divide and for an inclusive digital transformation is based on the examination above. Due to the focus and the capacity of our interdisciplinary research group on "Diversity and Inclusion in Digital Societies", a triangulation of the data for a quantitative and qualitative more detailed examination of the participants and the digital divide was essential. We therefore selected a Twitter/X corpus of 4 months between January, 1st 2023 and April, 18th 2023 consisting of 7,963 Tweets in German with **208,829 tokens.**

Since we were also interested in gaining a first overview of the actors participating in the Twitter discourse about inclusion and diversity, we collected all tweets from January to March 2023 and then drew a smaller sample of 700 tweets that generated at least 20 likes. After coding the actor type of the tweets' authors following parts of a codebook developed by Schmid-Petri et al. (2023), we removed all tweets coded as "unidentifiable actor". In doing so, our final sample consisted of 637 tweets (table 3):

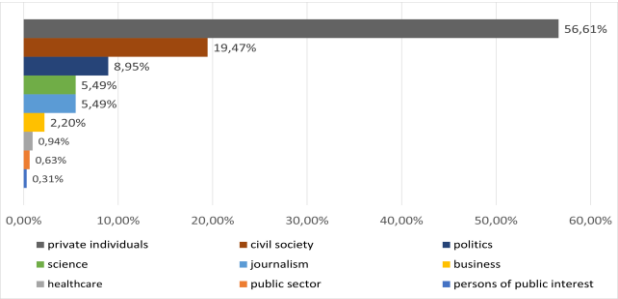
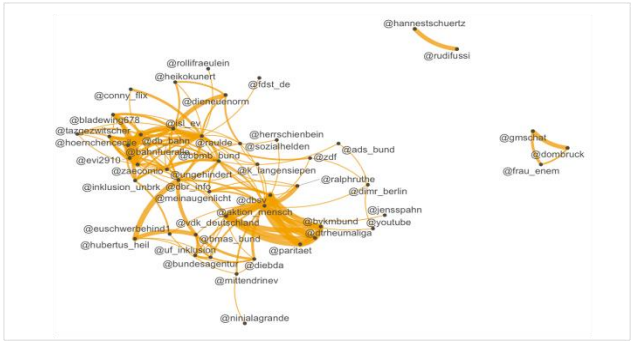


Table 3: Analysis of the participants

The study on participation reveals that the highest participation can be observed in the corpus by the group of "private individuals". The second highest participation is attributed to civic society. Both are indicators of community engagement among people interested in the topic of '(digital) accessibility`, 'inclusion', and 'disability' as well as organizations seeking inclusion for people with disability. In order to gain more insights into the participants, a further examination of participation was carried out by a second analysis from communication science. Opposed to this first limited study of the participants, we were able to include for the location of the main protagonists all 7963 Tweets – 208,829 tokens – in the time span of the three months between January and April 2023.

The table below shows the main participants with a broader outreach on social media.[7] A profile check of the most significant participants in figure 2 confirms that mainly individuals with a disability such as the disabled author, journalist, and activist Raúl Krauthausen but also the author Tanja Kollodzieyski a.k.a. 'Rollifreulein', and their organizations/self-representatives as well as in some cases even politicians participate in the discourse:

Figure 2: Most significant participants of the discourse



Further findings on disability engagement in the social media corpus is additionally generated by language-centered-data mining from the point of view of digital corpus linguistics, and in the service of human-centered-data-science. The frequency analysis of the vocabulary related to the different types of disability provides indications to the extent of digital activism of disabled individuals with different disability types. The study of the vocabulary reveals information on digital disability activism as follows:

(1) high participation of disabled people in the discourse
(2) higher frequencies of 'autism' & 'visual impairment'
(3) lower frequencies of 'mental (intellectual)'
(4) very low frequencies of 'psychic', and 'physical'
(5) extremely low frequencies of 'mental (emotional)' and 'learning'

The digital engagement in the discourse, of course, does not depend on the disability type but on the particular barriers people with disability face on social media. Due to a lack of image description, it is consequent that people with visual impairments face greater barriers on digital media than people with physical disabilities. **This outcome points out the need to consider inclusion according to the UN CRPD for aligning communication inclusively from the outset of the digital communication process.**

In our digital data between Jan. – Apr. 2023, over 700 out of 7963 examples prove the evidence for violating accessibility. We found (1) the only partial consideration of digital accessibility in digital communication of governmental & public stakeholders and the evidence (2) for the highest need to reduce digital barriers and transform official digital communication accessible to neurodiverse ('autism') & visually impaired individuals ('visual impairment'). Due to the lack of digital accessibility, people with disability organized their digital community and advocated for an inclusive official digital transformation by addressing the need for accessibility to (1) governmental organizations, (2) government spokespersons, (3) ministerial departments etc. by two characteristic types of interactions:

(1) comment under an exclusionary tweet
(2) retweet of an exclusionary tweet with a comment.

A detailed analysis of digital interactions confirmed that

[7] The research group „Diversity and Inclusion in Digital Societies" and the authors of this paper would like to thank Dr. Stefanie Walter (Technical University of Munich) for supporting this study by providing this visualization.

people with disability transformed digital communication towards more accessibility as some individuals and offices reacted to the criticism due to the lack of digital accessibility. In this way, people with disabilities made a major contribution to a more inclusive digital transformation.
Based on the findings of the linguistic analysis, a review of legal regulations for digital accessibility in Bavaria, Germany, and the EU was carried out. The legal review shows that there are relevant legal regulations for digital accessibility at all three organizational levels, but their implementation must be more strictly observed and monitored.

5. Conclusion

Our interdisciplinary study in the service of Human-Centered-Data-Science highlighted first the raising awareness for inclusion and (digital) accessibility on social media as well as on traditional media digitally in progress. Furthermore, our research points out the opportunities social media offer to address inclusion and (digital) accessibility and question the non-inclusive communication on the Internet. On the one hand, computer-mediated-communication is consequently used for boundary-drawing processes between a peer community of people with disabilities and their institutional representatives on social media, on the other hand for gaining more public attention for inclusion and accessibility. In this respect, the linguistic corpus analyses referred to various negotiations of social participation in the digital space, which leads to corresponding actions of the disability peer community for an inclusive digital transformation to enlarge accessible communication within and by using social media. The simultaneous evaluation of the outcome of the linguistic and legal analysis leads to our recommendation for processes of public digital communication to public stakeholders. We stated that digital accessibility goes beyond legislative regulations. Rather, it must also be taken into account and included in planning and realizing accessible digital public communication from the outset in line with all regulations for digital accessibility and with the UN Convention on the Rights of Persons with Disabilities. To sum up, digital accessibility needs to become part of the public digital communication process.

6. References

Artiles, A. J., Dyson, A. (2005). Inclusive Education in the Globalization Age: The Promise of Comparative Cultural Historical Analysis. In: D. Mitchell (Ed.), *Contextualizing Inclusive Education*. London: Routledge, pp. 37--62.

Bittlingmayer, U. H., Sahrai D. (2017). Inklusion als Anti-Diskriminierungsstrategie. In: A. Scherr, A. El-Mafaalani & G. Yüksel (Eds.), *Handbuch Diskriminierung. Springer Reference Sozialwissenschaften*. Wiesbaden: Springer VS, pp. 683--699.

Budde, J., Hummrich, M. (2015). Intersektionalität und reflexive Inklusion. In: *Sonderpädagogische Förderung*, Nr. 2. Weinheim: Beltz Juventa, pp. 165--175.

Durkheim, E. 1933. *The Division of Labor in Society* (first English translation by George Simpson). New York: Macmillan.

Fábián, A., Trost, I. (2023). Digital Corpus Linguistic Analysis of the Language of Inclusion, Discrimination and Exclusion of people with disability on social media. In: L. Cotgrove, L. Herzberg, H. Lüngen & I. Pisetta (Eds.): *Proceedings of the 10th International Conference on CMC and Social Media Corpora for the Humanities 2023*. Mannheim: Leibniz-Institut für Deutsche Sprache (IDS), pp. 65-72.

Goffman, E. (1961). *Asyle. Über die soziale Situation psychiatrischer Patienten und anderer Insassen*. Frankfurt a.M.: Suhrkamp.

Haage, A. (2020). Soziale Medien und Netzwerke. In: S. Hartwig (Ed.), *Behinderung. Kulturwissenschaftliches Handbuch*. Berlin: J.B. Metzler Verlag, pp. 289--296.

Herrera, L.C., Gjøsæter, T. (2022). Community Segmentation and Inclusive Social Media Listening. In: R. Grace & H. Baharmand (eds.), *ISCRAM 2022 Conference Proceedings – 19th International Conference on Information Systems for Crisis Response and Management*, 1012–1023. Tarbes, France. https://idl.iscram.org/files/luciacastroherrera/2022/2467_LuciaCastroHerrera+TerjeGjosaeter2022.pdf

Husemann, R. (2017). Einige Bemerkungen zur Diskussion um Inklusion und Exklusion in soziologischer Sicht. In: M. Gercke, S. Opalinski & T. Thonagel (Eds.), *Inklusive Bildung und gesellschaftliche Exklusion. Zusammenhänge – Widersprüche – Konsequenzen*. Wiesbaden: Springer VS, pp. 61--85.

Marshall, Th. H. (1964). *Class, Citizenship, and Social Development*. Garden City, New York: Doubleday.

Scherr, A. (2017). Soziologische Diskriminierungsforschung. In: *Handbuch Diskriminierung, Springer Reference Sozialwissenschaften*. Wiesbaden: Springer VS, 39-58.

Schmid-Petri, H., Bürger, M., Schlögl, S., Schwind, M., Mitrović, J., & Kühn, R. (2023). The Multilingual Twitter Discourse on Vaccination in Germany During the Covid-19 Pandemic. *Media and Communication*, 11(1), pp. 293--305.

Schütz, A. (2011 {1957}). Gleichheit und Sinnstruktur der sozialen Welt. In: A. Schütz (Ed*.), Relevanz und Handeln 2*. Konstanz: UVK, pp. 171--250.

Stichweh, R. (2009). Leitgesichtspunkte einer Soziologie der Inklusion und Exklusion. In: R. Stichweh & P. Windolf (Eds.), *Inklusion und Exklusion. Analysen zur Sozialstruktur und sozialen Ungleichheit*. Wiesbaden: VS Verl. für Sozialwiss, pp. 29--42.

UN-Behindertenrechtskonvention (2013). *Inklusion. UN-Behindertenrechtskonvention*. Source: http://www.behindertenrechtskonvention.info/inklusion-3693/ [latest access: 26.04.2024].

Weber, M. (2005). *Wirtschaft und Gesellschaf: Grundriss der verstehenden Soziologie (zwei Teile in einem Band)*. Frankfurt a. M.: Zweitausendeins.

Zelena, A. (2020). The Psychology of Inclusion on New Media Platforms and the Online Communication. *Acta Universitatis Sapientiae, Communicatio* 7(1), pp. 54--67.

Making suggestions in students' forum discussions

Francisco Javier Fernández Polo
Department of English and German Studies
University of Santiago de Compostela
xabier.fernandez@usc.es

Abstract: In this paper, I delve into the nature of peer advice in undergraduate students' online discussions. I provide a corpus-based analysis of the internal structure and relative directness of suggestions in forum discussions where they offer feedback on a classmate's work. Previous studies have highlighted the intricate advising sequence, where advice is often accompanied by supportive moves. Additionally, advice can manifest in various forms, each conveying different levels of imposition. Students in the sample predominantly use indirect ways to formulate suggestions in the forums. Though non-directiveness is associated with peer advice, overly contained formulations, it is claimed, may weaken the strength of their proposals. Students often fail to ground their advice on specific issues and omit providing reasons for their suggested solutions, which may also compromise persuasiveness. Some of the findings may reveal signs of interlanguage in the way students frame suggestions.

Keywords: forum discussions, suggestions, peer-advice, academic English, English as a lingua franca

1. Introduction[1]

Giving advice plays a major role in our everyday lives and is crucial in the educational setting. Peer-advice, in particular, fulfils an increasingly important function in tertiary education and deserves attention from the scientific community (Locher & Limberg, 2012; Waring, 2012). The aim of this paper is to describe the internal composition and relative directness of the suggestions made by a group of students giving feedback on a classmate's work in forum discussions. Existing research has established the relative complexity of the advising sequence, where a central component containing the advice itself is often accompanied by a series of supportive moves (Suzuki, 2008), as well as the fact that advice may take various forms of realisation (Harrison & Barlow, 2009; Li, 2010; Locher, 2013) conveying different degrees of imposition. Given the intrinsic face-threatening nature of advice, a lot of this literature has also explored the different strategies employed by interactants to mitigate the imposition of the advisory act (DeCapua & Huber, 1995). As far as peer advice in particular is concerned, previous work has shown that it tends to be grounded, so as to justify the advice, to be mostly indirect and frequently accompanied by bonding work to mitigate the intrinsic face-threat (Morrow, 2006). Several studies have also demonstrated the existence of differences in the way suggestions are framed across cultures (Chentsova-Dutton & Vaughn, 2012) and the problems they pose for English as a Second Language (ESL) learners. De Capua & Dunham (2007), for instance, note that ESL students tend to overuse forms of making suggestions that convey high imposition, compared to native speakers. In the present corpus-based study, I seek to explore the following questions in relation to students' advice-giving practices in discussion forums: What types of advising sequences are found in the data? How are they internally organized in terms of discursive moves? How do students cater for the need to be formulate clear and convincing propositions, while, at the same time, managing to reinforce social links with their classmates? What interlanguage features, if any, are observed in the analysed materials and how could they compromise the effectiveness of the feedback and, eventually, of the learning experience?

2. Materials and methods

Materials come from the SUNCODAC corpus[2] (Cal Varela & Fernández Polo, 2020) and consist of all the posts submitted by 12 students to the forums organized over a semester-long undergraduate course on English into Spanish translation. The data constitute a stratified sample of the forum participants, with a balanced representation of posts by gender (male/female) and student's final grade in the module (excellent/medium). The materials are taken from the English sub-component of SUNCODAC. In the forums, participants collaborate in the construction of an optimal translation. One of the students shares a first draft and receives feedback from a group of classmates. The uploaded translation functions as a sort of advice-seeking move, and the feedback posts as the advice-giving. This interpretation is supported by the type of language used by participants themselves to describe their contributions, where the word "suggestion" is frequently employed. Participants in the forums include students from different linguo-cultural backgrounds, "local" Spanish students and exchange students from different backgrounds, mostly European, but also a large contingent of Chinese students. English, the mandatory language of participation in the forums, functions as a lingua franca for participants. The 12 students in the sample are all "local students", native speakers of Spanish or Galician. A total of 384 individual advising sequences, our unit of analysis, were identified in the data and closely examined to establish their move

[1] This research was funded by the Spanish Ministry of Science and Innovation (co-founded by FEDER), grant nr.: PID2021-122267NB-I00.

[2] The corpus is freely accessible and can be consulted at http://www.suncodac.com/, where you can also find all relevant information on the search interface and possibilities, as well as on the data collection and preparation procedures, current holdings, etc.

structure (Suzuki, 2008), relative directness (Borderia-Garcia, 2006; Locher, 2013; Martínez Flor, 2005) and presence of various face-saving strategies. The results were registered in an Excel data sheet to systematize the description and facilitate the identification of patterns in the data and their eventual quantification. The identification of moves was arrived at inductively through several rounds of close analysis and inspired by previous work on advice-giving in other contexts (see above).

3. Results and discussion

As regards their internal structure, the advising sequences in our data consist of one or a combination of the following major moves: the Suggestion itself, with the author's specific proposal to improve the translation; Problem, identifying and describing, more or less explicitly, what aspect of the translation needs improving/changing; Justification, with the reasons adduced by the author for his/her proposal; Finality, describing how the translation will be improved by implementing the author's suggestion; References, to dictionaries, websites, etc., to reinforce the credibility of the proposal. Table 1 contains the raw frequencies of each of these moves in the sample. Other minor moves, which are exceptionally found in the data, include Mitigation, to redress the face-threatening effect of the suggestion, and which can be Preventative (PM) (before the suggestion) or Remedial (RM) (once the suggestion has been made, e.g., *I wouldn't change the name of Soren's collection, I am in favour of maintaining proper names, but it is something personal, it is all about likes and dislikes*).

Type of move	**N**
Suggestion	374
Problem	124
Justification	144
Finality	27
References	78

Table 1. Raw frequencies of major advising moves in the sample.

The following example, containing up to 5 different moves, illustrates a maximum degree of elaboration, underscoring the author's effort to explicate the nature of the problem and provide evidence supporting the proposal, while, at the same time, taking into account the need to preserve the relationship with the addressee.

> (PM) All the words in italics and the collocations are really accurate. (P) I just find one thing which in my opinion sounds a bit despective or agressive in a way. Where you translated "idílicas tardes sin hacer absolutamente nada." it sounds to me as if you were saying they are lazy or spending time in unprofitable things . (J) What I understood from the original text is that you can enjoy that other part after meals, having fun and talking to the rest of the people in the table , without worrying or hurrying . (S) So I chose "idílicas tardes de ocio sin fin" or "idílicas tardes de recreamiento" (R) as you can see in the second entry of the DEL, http://del.rae.es/?id=VVjjOMS recrear buscarconj . gif Del lat . Recreāre . 1 . tr . Crear o producir de nuevo algo . 2 . tr . Divertir , alegrar o deleitar . U . t . c . prnl . (sic)

All the sequences contain at least one P and/or one S. As shown elsewhere (Suzuki, 2008), P can be the sole component in the advising sequence, with the suggestion being left for the reader to infer, e.g., *With regards to the verb "caber" I agree with 16JLL, the structure in which it was inserted is not grammatical in Spanish.* The absence of an S move, however, is an option that is extremely rare in the data (2.6%).

Both P and S can also be simple (P, S) or complex (PC, SC), when S is followed by a specification/elaboration of the advice, e.g., *we have to make an aproximative measurement: in the first case walls 12 feet thick I would tranlate something like "muros de más de tres metros de espesor"*; or P is complemented by a more precise description of the identified issue, e.g., *with "contemplar arte" I have some problems. (...) i think is not accurate enough for this colloquial context*. Figures for both PC (n=9) and SC (n=30) are rather low in the sample, less than 10% of their respective categories (see Table 1).

When P and S combine in a single advising sequence, there are basically two possibilities regarding their relative position in the sequence (Waring, 2012): the problem may precede the advice, which is the “default” order, or the problem may follow the advice. In the sample, P+S structures (n=90) far outnumber S+P (n=25).

Both P and S can be more or less direct, and various mitigating strategies may be employed to soften either criticism or advice (see Table 2). Following Borderia-Garcia (2006), in terms of directness, S can be broadly classified into *Direct advice*, where the directive force is clearly signaled (performatives, imperatives, deontic modals, etc.), *Indirect advice*, where some sort of mitigation is performed (hedges, participant shift, impersonal sentences or passive voice, where mitigation operates by virtue of the addressee not being explicitly mentioned, etc.) and *Non-conventional indirect advice*, where the advice is merely hinted at, e.g. *As it has been mentioned, ‘dar’ sounds more informal than ‘entregar’* (implying: *I suggest you use "entregar" instead of "dar",* or just the opposite, because the sequence is actually ambiguous).

Type of suggestion	**N**
Performative	5

Imperative	0
Deontic modal	50
Participant shift	183
Question	2
Presenting the suggestion as others' opinions	105
Impersonal sentences	23
Inclusive WE	33
Passive voice	13
Hints	128

Table 2. Raw frequencies of level of directness of S. Colour shades indicate intensity of directness.

Direct advice forms are notably rare in our data, except for modal verb directives, and of moderate strength at that, with the more forceful *have to/must* only accounting for just about 1 in 5 modals. Moreover, modal-directive suggestions tend to be heavily mitigated in the data, to minimize imposition, through the use of inclusive pronouns (e.g., *we should avoid repetition*), hedges (*I think you could try to use other options*, etc.), the passive voice (*should be translated as*), impersonal structures (*'Los MacCarthy' should go without the final 's*), or by presenting the suggestion as someone else's opinion (*As some other people said, I think that you should use italics with the words "in situ"*).

Indirect advice is most frequently performed by means of a "participant shift", mostly consisting in what is generally known as "suggestions by example" (West, cited in Waring, 2012), e.g. *I chose the verb "dio" instead of "entregó"*. Suggesting a particular course of action by giving an account of personal experience in a similar situation is a mild form of advice, which has often been associated with peer-advice (Kouper, 2010). Finally, other major indirect ways of giving advice in English, which are widely practiced in the ELT classroom, are notoriously underrepresented in the analyzed materials; for instance, formulating the suggestion as a question, a popular formula in native English, only occurs twice in the data.

4. Conclusion

The present study has explored the type of advice-giving strategies employed by a small group of undergraduate students participating in online forum discussions. The results show, in general, a marked preference for rather indirect forms of advice, with *Indirect suggestions* and *Hints* far outnumbering direct suggestions in the data. Indirectness may be the participating students' response to the perceived threat to the forum moderator's face posed by the negative assessment of the moderator's proposals and the intrinsic imposition of the suggested improvements. Indirectness would be the students' attempt to redress the implicit threat to the moderator's face inherent to the advice-giving situation. Preference for non-directiveness seems to be a characteristic feature of advice-giving situations (Locher, 2013), particularly so of peer-advice where it is important to downplay the implied adviser's superiority over the advisee. Peer-tutoring situations are one such context where non-directiveness is expected (Waring, 2012).

The "relational" benefits that students derive from making less prescriptive and more "contained" suggestions might be outweighed by a loss of persuasiveness, which is to be added to other problems observed in the construction of the students' advice. Two key moves in the advising sequence, P and J, register rather low figures in the data. In other words, students very often omit to ground their advice on a clearly identified issue in the moderators' translation and fail to provide reasons for their suggested improvements, two rhetorical moves that are central to the advising sequence.

One of the characteristics of advice-giving in peer-tutoring (Waring, 2012) is a relative elaborateness of the advising move itself, where a general description of a problem or a suggestion is frequently followed by an "expansion", presenting the idea in more concrete or detailed terms. What is more important, as stated by Waring (2012, p. 106) "One way to gauge the helpfulness of particular advising, for example, may hinge upon whether the advice comes with specific suggestions of implementation." Instances of elaborate problems and suggestions are relatively rare in our data, which adds up to the relative vagueness of some of the formulations, e.g. *does not sound good*; *does not sound very clear*; *looks a bit strange*, etc. leave the reader wondering about the nature of the problem. Such inexplicitness might be more characteristic of the advising sequences in the medium-grade students' posts, but the analysis of differences in the way various groups of students construct their suggestions is beyond the scope of the present study.

There are signs of a possible interference from the students' mother tongue, e.g., the extremely low number of question-suggestions in the students' texts may be motivated by the fact that this structure is rarely used to make suggestions in the students' native language, where such negative politeness strategies are less popular than in English (Hickey, 1991). There are other characteristic interlanguage features in the data, such as the finding that the most frequent modal verb realizing S is *should*, confirming De Capua & Dunham's (2007) observation of a "tendency of ESL/EFL grammar texts to associate *should (not)* with the giving of advice" (p. 326). The SUNCODAC corpus has a huge potential as a pedagogic tool in the English for Academic Purposes classroom (Cal Varela & Fernández

Polo, 2024). In it, students may find inspiration for their writing from native speaker materials and student models – what to include in their posts and how to write specific sections. Teachers may find authentic student production samples for class discussion of good and bad practice, as well as native and learner materials for comparison, which may be eventually used to improve future student participation in forum discussions.

.

5. References

Borderia-Garcia, A. M. (2006). The Acquisition of Pragmatics in Spanish as a Foreign Language: Interpreting and Giving Advice . Ph.D Thesis. University of Iowa.

Cal Varela, M., & Fernández Polo, F. J. (2020). Suncodac. A corpus of online forums in higher education. *Nexus*, *2020*(02), 44–52.

Cal Varela, Mario, and Francisco Javier Fernández Polo. 2024. "Using the SUNCODAC corpus to teach effective communicative skills in student forum discussions." Paper presented at the *II Congreso Internacional de Aplicacións da lingüística de corpus na didáctica das linguas*. Santiago de Compostela, 13-14 May, 2024.

Chentsova-Dutton, Y. E., & Vaughn, A. (2012). Let me tell you what to do: Cultural differences in advice-giving. *Journal of Cross-Cultural Psychology*, *43*(5), 687–703. h

DeCapua, A., & Dunham, J. F. (2007). The pragmatics of advice giving: Cross-cultural perspectives. *Intercultural Pragmatics*, *4*(3). pp. 319-342.

DeCapua, A., & Huber, L. (1995). "If I were you...": Advice in American English. *Multilingua - Journal of Cross-Cultural and Interlanguage Communication*, *14*(2), pp. 117–131.

Harrison, S., & Barlow, J. (2009). Politeness strategies and advice-giving in an online arthritis workshop. *Journal of Politeness Research*, *5*(1), 93–111.

Hickey, L. (1991). Comparatively polite people in Spain and Britain. *Association for Contemporary Iberian Studies, 4*(2), pp. 2-7.

Kouper, I. (2010). The pragmatics of peer advice in a LiveJournal community. In *Language@Internet 7*, p. 1). www.languageatinternet.de,

Li, E. S. (2010). Making suggestions: A contrastive study of young Hong Kong and Australian students. *Journal of Pragmatics*, *42*(3), 598–616.

Locher, A. M., & Limberg, H. (2012). Introduction to advice in discourse. In H. Limberg & M. A. Locher (Eds.), *Advice in discourse* (pp. 1–27). John Benjamins Pub. Co.

Locher, M. A. (2013). Internet advice. In S. Herring, D. Stein, & T. Virtanen (Eds.), *Pragmatics of Computer-Mediated Communication* (pp. 339–362). de Gruyter.

Martínez Flor, A. (2005). A theoretical review of the speech act of suggesting: Towards a taxonomy for its use in FLT. *Revista Alicantina de Estudios Ingleses*, *18*, pp. 167–187.

Morrow, P. R. (2006). Telling about problems and giving advice in an Internet discussion forum: Some discourse features. *Discourse Studies*, *8*(4), pp. 531–548.

Suzuki, T. (2008). *A Study of Lexicogrammatical and Discourse Strategies for 'Suggestion' with the Use of the English Speech Act Corpus(1).* JACET (Japan Association of College English Teachers) Annual Convention 2008, Waseda University.

Waring, H. Z. (2012). The advising sequence and its preference structures in graduate peer tutoring at an American university. In H. Limberg & M. A. Locher (Eds.), *Advice in Discourse,* pp. 97–118). John Benjamins Publishing Company.

Analysis of Socially Unacceptable Discourse with Zero-shot Learning

Mohamed Rayane GHILENE, Dimitra NIAOURI, Michele LINARDI, Julien LONGHI
ENSEA Engineering School, ETIS UMR-8051 CY Cergy Paris Université, AGORA CY Cergy Paris Université
rayane.ghilene@ensea.fr, {michele.linardi, dimitra.niaouri, julien.longhi} @cyu.fr

Abstract

Socially Unacceptable Discourse (SUD) analysis is crucial for maintaining online positive environments. We investigate the effectiveness of Entailment-based zero-shot text classification (unsupervised method) for SUD detection and characterization by leveraging pre-trained transformer models and prompting techniques. The results demonstrate good generalization capabilities of these models to unseen data and highlight the promising nature of this approach for generating labeled datasets for the analysis and characterization of extremist narratives. The findings of this research contribute to the development of robust tools for studying SUD and promoting responsible communication online.

Keywords: Socially Unacceptable Discourse, Machine Learning, Weakly-Supervised Learning, Explainable Analysis

1. Introduction

Large Language Models (LLMs) have showcased remarkable capabilities in Natural Language Processing (NLP) thanks to their contextual understanding of word embeddings, which have proven to be useful in multiple tasks, including text answering, text generation, and data annotation. LLMs have also shown potential for text classification tasks such as sentiment analysis (Zhang et al., 2023a) by leveraging prompt learning.

In recent years, the spread of Socially Unacceptable Discourse (SUD), including hate speech and toxic comments, in various online platforms has underscored the need for novel tools able to identify and characterize these harmful discourses. However, developing robust automatic SUD classifiers comes with multiple challenges. For instance, the challenge of adopting a universal definition of SUD due to the numerous discourse characterizations causes ambiguity and subjectivity in corpora adopted to train Machine Learning (ML) models (Kocon et al., 2021). Such a scenario poses significant challenges to the creation of well-annotated SUD text corpora that can extensively evaluate the quality of state-of-the-art classification models in large-scale scenarios.

SUD Classification challenges LLMs have obtained state-of-the-art performance in SUD text classification tasks. In this sense, Carneiro et al. (2023) have recently shown that Masked Language Models (MLM) represent a strong candidate classifier option in multiple online annotated corpora. At the same time, Causal Language Models (CLM), which are LLM variants specifically trained to learn cause-effect dynamics (usually adopted by generative AI) can also be successfully leveraged in hate speech classification (Zhang et al., 2023b).

Despite the effectiveness of these models, we note that LLMs lack generalizability in SUD modeling due to their nature, which consists of understanding statistical relationships between words rather than modeling the meaning of these words within their context. Zhang et al. (2023) show that LLMs often obtain solid classification performance in the presence of language stereotypes (e.g., race or religion-related).

On the other hand, in a large-scale context (Carneiro et al., 2023), where heterogeneous subdomains of toxic speech require to be differentiated (i.e., multi-class classification) LLMs are not capable of providing accurate classification due to the presence of overlapping characteristics among different speech classes, but also for the presence of subtle linguistic nuances that require to understand the underlying context to be detected.

Moreover, the annotation schema plays a crucial role in the supervised model training. Often, SUD annotation is subjective and prone to biases resulting from the annotator's background, gender, first language, age, and education (Al Kuwatly et al., 2020). For instance, significant disagreement among annotators from different cultures regarding the offensiveness of online language has been reported in previous studies (Thorn Jakobsen et al., 2022).

Contribution In this work, we present a novel SUD analysis framework, in which we adopt a zero-shot learning paradigm for the automatic detection and characterization of SUD in a large-scale context composed of multiple heterogeneous corpora. Specifically, we leverage natural language inference (NLI) pre-trained models to perform SUD inference (a.k.a. entailment) in text instances. The benefit of this approach is two-fold: first, we do not require data complying with a fixed annotation schema, which may be prone to human bias, second, it will permit to leverage human expertise for hypothesis engineering and validation (Goldzycher and Schneider, 2022), where the users can incorporate their understanding of a specific domain or field to guide the classification process.

2. SUD Framework based on Natural Language Inference

In our solution, we leverage Natural language Inference (NLI) pre-trained models, which are a specific type of NLP models trained to understand the relationship between two pieces of text, namely the *premise* and the *hypothesis* (a new text, potentially related to the premise).

2.1. Entailment Template

To define premise-hypothesis entailment, we follow a methodology similar to the one proposed for text classifi-

Premise (t)	Hypotheses	Candidate Labels	Entailment Score
what's the difference between a pencil arguing and a woman arguing a pencil has a point	This example is	hate	**0.43**
	This example is	offensive	0.35
	This example is	toxic	0.22

Table 1: Entailment-based zero-shot classification. For every text t (premise) in the dataset, we create multiple hypothesis by considering several known SUD labels.

cation (Gera et al., 2022), adapted to perform unsupervised data labeling.
In this regard, we showcase an illustrative example, drawing inspiration from prior research (Yin et al., 2019) which we have tailored to SUD analysis, as depicted in Figure 1. Here, a hateful premise can be assigned to different labels (hypothesis) according to the perspective under the lens (sentiment, tone of the speech, topics, etc.).

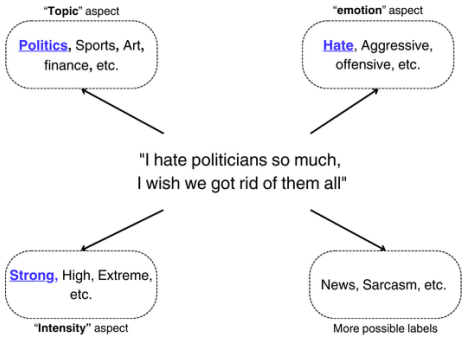


Figure 1: A piece of text can be assigned labels that describe the different aspects of the text. Relevant labels are in blue. Different characterizations of a hateful stance are at the basis of hate speech analysis (Qian et al., 2019).

We thus propose an entailment-based framework, where we couple each premise (text item in a corpus) with a hypothesis stating which class it belongs to. We construct a pair (text item/hypothesis) for each possible SUD class present in the annotation schema of the dataset. Constructed pairs become the input of an NLI model that infers a confidence (entailment) score. In this respect, we consider the output of the softmax layer [1] of an NLI model, where for each hypothesis a probability is assigned between 0 (contradictory hypothesis) and 1.0 (entailed hypothesis).
In Table 1 we report an entailment example that we obtain using a zero-shot learning paradigm to perform an unsupervised premise/hypothesis entailment. Note that a hypothesis is composed of a prefix and a candidate label arbitrarily chosen by the user.

2.2. Entailment Models

To perform zero-shot entailment-based text classification on the SUD data, we use models trained specifically for natural language interference (NLI). Such models are pretrained on the MNLI (Multi-Genre Natural Language Inference) dataset (Williams et al., 2018) which is a large collection of sentence pairs used to evaluate models on their ability to understand entailment between sentences.
It contains over 433,000 sentence pairs in English, drawn from ten different genres of written and spoken text, including news articles, fiction, and conversations. Each pair consists of a premise sentence (source) and a hypothesis sentence (target).
Models trained on the MNLI dataset have the ability to generalize well to different types of textual data, thanks to the diversity of genres they have encountered in the training procedure. For the SUD classification task, we use the following models:

- **Roberta-large-mnli**, BERT (Devlin et al., 2019), which is a transformer-based language model pretrained on English text using a masked language modeling (MLM) objective and fine-tuned on the Multi-Genre Natural Language Inference (MNLI) corpus.

- **Bart-large-mnli** (Lewis et al., 2020), which is a transformer encoder-decoder (seq2seq) model with a bidirectional (BERT-like) encoder and an autoregressive (GPT-like) decoder. BART is pre-trained by corrupting text with an arbitrary noising function and learning a model to reconstruct the original text. In our work, we consider the model "facebook/bart-large-mnli", BART version pre-trained on MNLI (Williams et al., 2018) dataset, for Entailment-based Zero shot classification.

  We also consider models trained on other NLI datasets:

- **xlm-roberta-large-xnli-anli**, is a variant of the XLM-RoBERTa architecture proposed in (Conneau et al., 2020), fine-tuned on the XNLI (Cross-lingual Natural Language Inference) (Conneau et al., 2018) and ANLI (Adversarial Natural Language Inference) (Williams et al., 2020) datasets. Its primary application is in cross-lingual natural language inference, which involves determining the relationship (such as entailment, contradiction, or neutrality) between pairs of sentences across multiple languages.

- **MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7**, multilingual natural language inference (NLI) model based on the mDeBERTa-v3 architecture, fine-tuned on a combination of the XNLI dataset and an additional multilingual NLI dataset with 2.7 million examples. The mDeBERTa-v3 architecture enhances its performance by incorporating improvements in transformer design, such as disentangled attention and enhanced mask decoder.

3. Empirical Evaluation

To validate our solution, we perform zero-shot entailment-based classification on several publicly available datasets (Carneiro et al., 2023). Below, we introduce the datasets employed and the results acquired. For the sake of repro-

[1] In our case, the softmax layer takes a textual feature vector (learned by the model) of real-valued numbers, transforming it into a probability distribution over a set of possible categories (hypothesis).

Dataset	Source	Sample type	# Samples	Labels
Davidson	(Davidson et al., 2017)	Tweets	25,000	hate, offensive, neither
Founta	(Founta et al., 2018)	Tweets	100,000	abusive, hate, neither
Fox	(Gao and Huang, 2017)	Threads	1,528	hate, neither
Gab	(Qian et al., 2019)	Posts	34,000	hate, neither
Grimminger	(Grimminger and Klinger, 2021)	Tweets	3,000	hate, neither
HASOC2019	(Mandl et al., 2019)	Facebook, Twitter posts	12,000	hate, offensive, profane, neither
HASOC2020	(Mandl et al., 2020)	Facebook posts	12,000	hate, offensive, profane, neither
Hateval	(Basile et al., 2019)	Tweets	13,000	hate, neither
Olid	(Zampieri et al., 2019)	Tweets	14,000	offensive, neither
Reddit	(Yuan and Rizoiu, 2022)	Posts	22,000	hate, neither
Stormfront	(De Gibert et al., 2018)	Threads	10,500	hate, neither
Trac	(Kumar et al., 2018)	Facebook posts	15,000	aggressive, neither

Table 2: Summary of datasets (Carneiro et al., 2023)

ducibility, the implemented source code used in the evaluation is publicly available on a public repository [2].

3.1. Datasets

We conducted our evaluation in 12 publicly available datasets containing up to 12 different classes of SUD (Carneiro et al., 2023). In Table 2 we report a detailed overview of the English datasets considered in our study.

3.2. Evaluation of SUD Classifiers

The first goal of our evaluation is to compare the entailment models (unsupervised) with the results we obtain adopting a supervised classifier that has been specifically trained over the annotation schema provided in each dataset.

Such experiment will permit us to answer the question: ***How performance of an elastic and unsupervised method that does not rely on prior SUD knowledge (i.e. the entailment-based zero-shot learning) compare to the ones of a classifier trained over SUD knowledge?*** For this latter, we consider a state-of-the-art MLM, namely BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019)

Note that Masked language models (MLMs), like BERT, are DL models trained to reconstruct masked tokens within the input sequence. Specifically, MLMs learn to predict the original vocabulary identity of a masked word, relying solely on its contextual cues. The significant advantage of those models lies in their bidirectional context, considering both preceding and subsequent tokens during the prediction process. In this work, we fine-tune BERT with the available SUD labels in each evaluated dataset.

We also consider a *shallow* learning baseline such as Logistic Regression (LR) (Grimm and Yarnold, 1995) applied to the numerical representation of tokenized text (text vectorization) [3].

Note that the entailment models we adopt were not pretrained by leveraging any available ground truth of SUD, and thus they are unsupervised methods in that respect.

[2] https://github.com/rayaneghilene/ARENAS_Automatic_Extremist_Analysis/tree/main/Entailment_framework

[3] https://keras.io/api/layers/preprocessing_layers/text/text_vectorization/

Hypothesis Testing	roBERTa	BART	mDeBERTa	XLM-roBERTa
this text contains {} speech.	**45.7**	27.6	27.7	**40.9**
this text conveys {} speech.	40.8	34.7	27.6	35.8
this text reflects {} speech.	38.3	35.5	33.8	33.8
this text shows {} speech.	35.1	38.5	24.6	35.7
this text implies {} speech.	33.2	39.6	28.4	32.1
this text reveals {} speech.	37.8	41.6	26.2	32.8
this text exhibits {} speech.	38.8	33.3	21.2	40.4
this text portrays {} speech.	33	36.3	31.1	31.6
this text discusses {} speech.	34.8	37.9	34.3	34.5
this text addresses {} speech.	34.2	38	35	37.1
this text illustrates {} speech.	35.9	43	30.7	32.2
this text expresses {} speech.	44.5	35.7	**35.6**	32.9
this text articulates {} speech.	45.1	42.5	33.2	31
this text suggests {} speech.	30.1	38.6	27.8	32.8
this text narrates {} speech.	43.2	40.5	34.5	35.1
this text questions {} speech.	32.6	42	6.9	28.6
this text demonstrates {} speech.	35	42.2	22.9	31.5
this text supports {} speech.	22.6	**44,4**	32	31.9
this text has {} speech.	41.1	32.5	10.6	39.3

Table 3: Hypothesis Testing F1 Scores

We base our comparison on the macro F1 score, which is an averaging method for the F1 score that's recommended when working with class imbalance. F1 score is a harmonic mean that combines two performance measures for text classifiers: precision (P) and recall (R). These metrics are computed as follows: $\mathbf{R} = \frac{\text{True Positives}}{\text{True Positives+False Negatives}}$ and $\mathbf{P} = \frac{\text{True Positives}}{\text{True Positives+False Positives}}$.

The F1 score is calculated based on these metrics as $\textbf{F1 Score} = 2 \cdot \frac{\text{P·R}}{\text{P+R}}$. And then the macro F1 score is computed as $\textbf{Macro F1} = \frac{1}{C} \cdot \sum_{i=1}^{C}(F1_i)$, where C is the total number of classes.

3.2.1. Template Selection

In our evaluation, we note that hypothesis construction plays a crucial role in NLI model performance, which has a sensitive and different impact on the considered NLI models, each adopting a different Token masking procedure at the pre-training stage.

In Table 3, we report the hypothesis templates we consider in our work. In detail, we have tested different active parts, i.e., the verb in the formulation, noticing a remarkable impact (+/- 20 in average F1 score) on average SUD classification performance that we report for each model in Table 3. We observe that the four considered NLI models reach the best F1 score using three different hypothesis templates, which we use in the remaining part of the evaluation.

In the same manner, we note that using the word *neither* in the hypothesis template does not provide any contextual in-

	Supervised SUD classification		Unsupervised SUD classification (entailment-based)			
Dataset	BERT	LR	Bart-large-mnli	Roberta-large-mnli	xlm-roBERTa	mDeBERTa
Davidson	73	69.5	47.3	44.7	41.5	19.1
Founta	70.1	73.7	57.4	57.5	42.8	30.5
Fox	47.8	69.7	56.1	55.2	52.5	40.9
Gab	87.5	89.0	64.7	67.1	58.3	47
Grimminger	51.9	50.4	**52.5**	**56.1**	48.8	30.9
HASOC2019	32.9	39.9	27.5	30.9	17.8	17.2
HASOC2020	41.7	52.5	36.7	42.7	20.4	15.4
Hateval	63.6	70.6	59.7	61.4	57.2	49.3
Olid	65.6	71.9	61.6	61.5	52.1	51.2
Reddit	81.7	83.0	56.3	58	50.9	38.1
Stormfront	66.9	68.4	62	62.6	55.2	40.2
Trac	67.1	69.2	52.1	64.2	61.7	48.2

Table 4: Macro F1 Score (%) of supervised SUD classification VS Entailment-based unsupervised SUD classification with the NLI models.

formation to the inference phase of the neutral class, resulting in sensibly low classification performance. We obtain the best performance using the term *neutral speech* in the hypothesis instead of the word *neither* found in each dataset annotation schema (see Table 2).

3.2.2. Results and Discussion

We report experimental results in Table 4. As expected, entailment-based model classification shows slightly lower performance when using entailment models compared to a pre-trained MLM. However, this is not the case for all the datasets, in the Grimminger dataset, our approach outperforms the supervised counterparts, showing a better ability in considering the discourse context at the entailment stage, rather than leveraging correlations among text items in the training set, as in the case of the supervised counterparts.

Furthermore, Roberta-large-mnli and Bart-large-mnli exhibit overall better performance than xlm-roBERTa and mDeBERTa, suggesting that pre-training over the MNLI dataset, which covers a wide range of different spoken and written text is a more suitable choice for SUD analysis.

It is also important to note that such results are similar to the ones obtained by (Gera et al., 2022) when performing zero-shot entailment on other types of text classification. To the best of our knowledge, we are the first to adopt such techniques in SUD analysis.

To conclude, we also observe that there is no clear winner among the supervised classifiers, and a simple Logistic Regression represents an effective solution in the majority of the datasets.

Mitigating biases in the classification To further reduce user bias that may occur in the definition of the hypothesis we adopt GloVe (Pennington et al., 2014) token masking.

This procedure consists of masking tokens highly correlated with the label used in the hypothesis, causing the models to rely on the context provided by the remaining part of the speech in the classification task.

For each text, we mask the tokens with the highest GloVe similarity to the class name following the idea proposed in (Gera et al., 2022).

For example, when classifying *offensive* SUD, the words correlated to offensive language will be masked in the text.

The experimental results reported in Table 5 show that the effect of token masking comes only with a slight performance decrease (in most datasets) compared to the results obtained by the entailment models off-the-shelf. Such results suggest how entailment-based SUD classification can not only leverage class stereotypes, but it can potentially leverage the remaining part of the speech.

Dataset	Bart-large-mnli	Bart-large-mnli + Mask	RoBERTa-large-mnli	RoBERTa-large-mnli + Mask
Davidson	47.3	40.3	44.7	42.5
Founta	57.4	53	57.5	49.8
Fox	56.1	55.5	55.2	57
Gab	64.7	61.4	67.1	66.6
Grimminger	52.5	50.5	56.1	56.4
HASOC2019	27.5	23.3	30.9	29.8
HASOC2020	36.7	28.6	42.7	37.3
Hateval	60.8	58.6	61.4	61.3
Olid	61.6	59.5	61.5	61.8
Reddit	56.3	53.6	58	59.8
Stormfront	62	59.1	62.6	62.6
Trac	52.1	47.6	64.2	63.4

Table 5: **Zero-shot text classification with token masking** For each zero-shot entailment model and dataset, we compare the macro F1 score of the off-the-shelf model to its score when performing token masking.

4. Conclusion and Future Work

This paper investigates the effectiveness of zero-shot entailment using NLI models for SUD classification.

Through preliminary experimentation, these models showcased generalization capabilities comparable with supervised counterparts. Such a scenario highlights the entailment-based model's potentiality to exploit contextual information in the text rather than learning intra-class correlation using a fixed annotation schema, which may be sensitive to stereotypes of certain kinds of SUD.

The preliminary results we obtained motivate several future work directions. First, we would like to explore how to effectively learn templates that allow linguists to use semantically richer and unstructured annotation schemes, also studying scalability issues and tradeoffs of large entailment hypothesis spaces. We believe that such capability can support supervised learning models currently adopted in SUD analysis to reduce the impact of annotator bias and

sensitivity to class stereotypes.

This result will be a valuable advance for the CMC corpora community and work in corpus linguistics, allowing synergies between AI and corpus linguistics researchers.

5. References

Al Kuwatly, H., Wich, M., and Groh, G. (2020). Identifying and measuring annotator bias based on annotators' demographic characteristics. In *Proceedings of the Fourth Workshop on Online Abuse and Harms*.

Basile, V., Bosco, C., Fersini, E., et al. (2019). Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in twitter. In *Proceedings of the 13th international workshop on semantic evaluation*.

Carneiro, B. M., Linardi, M., and Longhi, J. (2023). Studying socially unacceptable discourse classification (SUD) through different eyes: "are we on the same page ?". *CoRR*, abs/2308.04180.

Conneau, A., Lample, G., Rinott, R., Williams, A., Bowman, S. R., Schwenk, H., and Stoyanov, V. (2018). Xnli: Evaluating cross-lingual sentence representations. *arXiv preprint arXiv:1809.05053*.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., GuzmÃ¡n, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale.

Davidson, T., Warmsley, D., Macy, M. W., et al. (2017). Automated hate speech detection and the problem of offensive language. In *Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM*. AAAI Press.

De Gibert, O., Perez, N., García-Pablos, A., et al. (2018). Hate speech dataset from a white supremacy forum. *arXiv preprint arXiv:1809.04444*.

Devlin, J., Chang, M., Lee, K., et al. (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: NAACL-HLT*.

Founta, A., Djouvas, C., Chatzakou, D., et al. (2018). Large scale crowdsourcing and characterization of twitter abusive behavior. In *Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM*.

Gao, L. and Huang, R. (2017). Detecting online hate speech using context aware models. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP*.

Gera, A., Halfon, A., Shnarch, E., et al. (2022). Zero-shot text classification with self-training. In *Proceedings of EMNLP*.

Goldzycher, J. and Schneider, G. (2022). Hypothesis engineering for zero-shot hate speech detection. *arXiv preprint arXiv:2210.00910*.

Grimm, L. G. and Yarnold, P. R. (1995). *Reading and understanding multivariate statistics.* American psychological association.

Grimminger, L. and Klinger, R. (2021). Hate towards the political opponent: A twitter corpus study of the 2020 US elections on the basis of offensive speech and stance detection. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EACL 2021*.

Kocon, J., Figas, A., Gruza, M., et al. (2021). Offensive, aggressive, and hate speech analysis: From data-centric to human-centered approach. *Inf. Process. Manag.*, 58(5).

Kumar, R., Reganti, A. N., Bhatia, A., et al. (2018). Aggression-annotated corpus of hindi-english code-mixed data. *arXiv preprint arXiv:1803.09402*.

Lewis, M., Liu, Y., Goyal, N., et al. (2020). BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*.

Mandl, T., Modha, S., Majumder, P., et al. (2019). Overview of the hasoc track at fire 2019: Hate speech and offensive content identification in indo-european languages. In *Proceedings of the 11th annual meeting of the Forum for Information Retrieval Evaluation*.

Mandl, T., Modha, S., Kumar M, A., et al. (2020). Overview of the hasoc track at fire 2020: Hate speech and offensive language identification in tamil, malayalam, hindi, english and german. In *Proceedings of the 12th annual meeting of the forum for information retrieval evaluation*.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*.

Qian, J., Bethke, A., Liu, Y., et al. (2019). A benchmark dataset for learning to intervene in online hate speech. In *Proceedings of EMNLP-IJCNLP*.

Thorn Jakobsen, T. S., Barrett, M., Søgaard, A., et al. (2022). The sensitivity of annotator bias to task definitions in argument mining. In *Proceedings of the 16th Linguistic Annotation Workshop (LAW-XVI) within LREC2022*.

Williams, A., Nangia, N., and Bowman, S. R. (2018). A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: NAACL-HLT*.

Williams, A., Thrush, T., and Kiela, D. (2020). Anlizing the adversarial natural language inference dataset. *arXiv preprint arXiv:2010.12729*.

Yin, W., Hay, J., and Roth, D. (2019). Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. In *Proceedings of EMNLP-IJCNLP*.

Yuan, L. and Rizoiu, M. (2022). Detect hate speech in unseen domains using multi-task learning: A case study of political public figures. *CoRR*, abs/2208.10598.

Zampieri, M., Malmasi, S., Nakov, P., et al. (2019). Predicting the type and target of offensive posts in social media. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Zhang, P., Chai, T., and Xu, Y. (2023a). Adaptive prompt learning-based few-shot sentiment analysis. *Neural Process. Lett.*, 55(6).

Zhang, Z., Chen, J., and Yang, D. (2023b). Mitigating biases in hate speech detection from A causal perspective. In *Findings of the Association for Computational Linguistics: EMNLP*.

"I Think your Translation is Great, but I have some Suggestions that may Help you (or not)". The Use of Concessives as Politeness Devices in Asynchronous Online Discussion Forums

Elsa González Álvarez, Susana M. Doval Suárez
Universidade de Santiago de Compostela
E-mail: elsa.gonzalez@usc.es,
susanamaria.doval@usc.es

Abstract

This paper aims to provide a characterization of concessives in an asynchronous online discussion forum, and to show the relevance of politeness for this characterization. A sample of 165 concessive clauses headed by *but* (henceforth, butCs) was extracted from the English component of SUNCODAC (Santiago University Corpus of Discussions in Academic Contexts). First, we tested whether butCs follow the pattern described by Couper-Cuhlen and Thomson (2000: 38) and refined by Musi et al. (2018). Then, we explored the co-occurrence of butCs with different lexical features which have been reported to be important for this categorization (Musi et al. 2018), and whether their distribution varies depending on the proposition. Our findings reveal that only a small percentage of butCs contain no instances of these features and that their typical distribution directly mirrors the semantic and interactional function of each proposition. The results also indicate that addressing the description of concessives by referring to their role as politeness devices is especially relevant for the study of L2 CMC contexts which include assessment and evaluation of peers' writing.

Keywords: L2 writing, computer-mediated communication, politeness, concession.

1. Introduction

1.1. Politeness in CMC

This paper aims to provide a characterization of concessives in an asynchronous online discussion forum, and to show the relevance of politeness for this characterization. More specifically, we will explore how Spanish EFL learners participating in the discussions use concession in combination with other politeness strategies in a collaborative pedagogical context. Therefore, we are interested in issues of face (Goffman, 1967) and politeness (Brown and Levinson, 1978; 1987). In the list of potentially face-threatening acts (FTAs), Brown and Levinson's theory of politeness includes orders, requests, suggestions, advice, reminders, warnings, offers, promises or criticism (1987: 66-67). These speech acts can be mitigated by using positive and negative politeness strategies, depending on whether they are used to protect positive face (i.e., the universal desire to be appreciated and socially accepted) or to protect negative face (i.e., people's desire to preserve autonomy). Examples of positive politeness strategies include attending to the interlocutor's needs or wants, seeking agreement, softening disagreement, including the writer and the reader in the activity, and showing praise or appreciation, among others. Negative strategies, on the contrary, include being indirect, minimizing an imposition, apologizing, and impersonalizing a situation, among others (Schallert et al., 2009: 718).

Even though Brown and Levinson's work has remained influential over the years, it has been frequently challenged. Thus, considerable criticism has come from Watts (1992, 2003), Locher (2004), Locher and Watts (2005, 2008), who argue that Brown and Levinson's model is not "in fact a theory of politeness but rather a theory of facework" that fails to account for "those situations in which face-threat mitigation is not a priority", such as aggressive or impolite behaviour (Locher and Watts, 2005: 10). Focusing on the interpersonal dimensions of language used in interaction, they develop the concept of 'relational work', i.e. "the 'work' that individuals invest in negotiating relationships with others" (Locher and Watts, 2005: 10). It is important to remark that in their view, Brown and Levinson's concept of politeness can still be used, but it should be regarded as only a small part of relational work, which, in turn "comprises the entire continuum of verbal behaviour from direct, impolite, rude or aggressive interaction through to polite interaction" (Locher and Watts, 2005: 11).

From the point of view of politeness, the online medium has several peculiarities which inevitably shape CMC interactions. On the one hand, it imposes certain limitations which make participants reinforce the interpersonal links with their partners using markers of affection, interactive responses, and group cohesion expressions (Fernández-Polo and Cal-Varela, 2017). On the other hand, the lack of non-verbal clues increases the importance of using politeness to avoid misunderstandings, since FTAs such as "disagreements, criticisms, requests for information or help, and requests for clarification of a prior message" (Schallert et al., 2009: 715) are typical of CMC interactions (Herring, 2023).

The interest of our study stems precisely from the fact that politeness issues are particularly relevant for those interactions including assessment or evaluation of peers' L2 writing, as is the case with the discussion forums in this study, where most participants are L2-English learners (Cal-Varela and Fernández-Polo, 2019; Pyo and Lee, 2019). In these language learning contexts, where the emerging virtual communities have been found to promote interaction and diminish anxiety of communication (Deris et al., 2015: 79), the presence of FTAs also leads participants to soften their comments through mitigation strategies.

1.2. Concessives and Politeness

In addressing the description of concessives by referring to their role as politeness strategies in a pedagogical CMC setting, we intend to fill a gap in existing research. Although little research has been conducted on the role played by concession in CMC, a few exceptions can be

found. Tanskanen and Karhukorpi (2008), for instance, explore how participants in e-mail conversations use concessives to repair claims that may cause disagreement and Musi et al. (2018) explore the argumentative and persuasive value of concession in non-pedagogical discussion forums.

Outside the CMC context, the literature on concessives has referred only indirectly to their role as politeness strategies. Thus, Biber's (1988) multidimensional approach associates concessives with other mitigating devices such as hedges or downtoners. Other studies highlight the use of concession to increase the hearer's positive attitude towards the speaker's opinion (Mann and Thompson, 1988), since "recognizing the validity of the hearer's standpoint before expressing disagreement can avoid face-threatening acts and is perceived as reasonable by the hearer" (Couper-Kuhlen and Thompson, 2000: 381).

The concept of concession used here is based on Couper-Kuhlen and Thomson's (2000: 381) definition of concessives as three-part sequences in which:

- The first speaker makes a point (X).
- The second speaker concedes the validity of this point (X').
- The second speaker makes a potentially contrasting point which denies the expectations created by (Y).

We also draw on Musi et al. (2018), whose characterization of Argumentative Concessives (ACs) is also based on Couper-Kuhlen and Thomson (2000). According to these authors, at a semantic level, the conceding proposition (or Proposition A) of ACs typically includes agreement or a positive evaluation of the statement previously presented by the other speaker, while the denial-of-expectations proposition (or Proposition B) tends to include (mitigated) criticism. Additionally, these authors suggest that ACs typically co-occur with a number of linguistic features:

- Hedges[1].
- Positive and negative sentiment words (since ACs usually contain opinion on the other posts).
- First and second personal pronouns and adjectives (since ACs "dialogically point to the stance taken by the previous speaker" (Musi et al., 2018: 10)).

2. Method

2.1. Research questions

In order to describe how a specific type of concessive (i.e. butC) is used in combination with other politeness strategies in a pedagogical online discussion forum (Santiago University Corpus of Discussions in Academic Contexts; SUNCODAC, 2021), our study addresses the following research questions:

1. Do butCs in SUNCODAC follow the pattern described by Couper-Kuhlen and Thomson (2000) and refined by Musi et al. (2018)?
2. How frequently do butCs in SUNCODAC co-occur with the lexical features considered in Musi et al. (2018): boosters, hedges, positive and negative sentiment words, and first and second personal pronouns and adjectives?
3. What is the distribution of these lexical features in Propositions A and B of the concessive?

2.2. Data source

The corpus used in this study (SUNCODAC, 2021[2]) consists of student forum discussions gathered over a span of four years at the University of Santiago de Compostela (USC). These discussions were an integral part of an English-to-Spanish translation course designed for second-year undergraduates, primarily majoring in English at USC. The forum served as a supplementary tool alongside traditional face-to-face teaching, and students actively contributed at three distinct time points during the semester: the beginning, middle, and end.

SUNCODAC contains a representation of English, Spanish and Galician used as first (L1) and second (L2) languages by students of different nationalities. The subjects are L1 and L2 English speakers of several L1 backgrounds, mainly Spanish, Galician, English and Chinese, but this study concentrates on L1-Spanish participants' productions in L2 English.

A detailed description of the activity can be found in Cal-Varela and Fernández-Polo (2020: 46-47). Every week, a practical session was allocated for in-class discussions on a translation topic, followed by an online discussion. To facilitate this, distinct weekly forums were created within the Moodle platform. Each forum was overseen by a student assigned as the moderator. The activity unfolded through five stages:

- Lecturers' instructions. A single opening post by the lecturers including the source text, the moderator's name, basic instructions, and deadlines.
- Moderator's first translation.
- Peer feedback. This is the core of the discussion and consists of messages where the moderator's classmates make comments and suggestions for improvement and discuss the suitability of different translation solutions.
- Moderator's improved version and summary of discussion.
- Instructor's assessment and appraisal of the activity

2.3. Procedure

A sample of 165 concessives headed by *but* (henceforth butCs) produced by L1-Spanish speakers was extracted from the English component of SUNCODAC, using the corpus search tool. This sample represents 15% of the overall occurrence of this marker in the whole corpus. The butCs are uniformly distributed across sections, gender

[1] Hedges can be defined as lexical and syntactic means of decreasing the writer's responsibility "for the extent and the truth-value of propositions and claims, displaying hesitation, uncertainty, indirectness, and/or politeness to reduce the imposition on the reader" (Hinkel, 2005: 30).

[2] http://www.suncodac.com/

groups and periods, i.e. we selected equal numbers of butCs for each level of the different variables used as corpus design criteria: gender, post section and post period.

The decision to include only butCs was motivated by the fact that this connective has been found to be the most frequent concessive marker in different discourse types[3] (Grote et al., 1997; Izutsu, 2008; Taboada and Gómez-González, 2012). Additionally, *but* represents 85% of concessive markers in discussion forums (Musi et al., 2018) and 52% of all concessive markers in the English component of SUNCODAC (Author 2021). Barth (2000: 418) explains that the reason for this prevalence of *but* is the fact that but-constructions "provide an opportunity for face work by leaving the speaker room to manoeuvre and by attending to the recipient's need for politeness".

The creation of the subcorpus was followed by the automatic extraction of examples containing the different lexical features under study using Wordsmith Tools 7 (Scott 2016), followed by manual disambiguation of examples. The list of lexical features was constructed by referring to previous studies. Thus, we used the lists of hedges appearing in Hyland (2005); the list of intensifiers used by Hinkel (2005); and, in order to select the positive and negative sentiment words, we chose the sentiment/opinion lexicon published by Hu and Liu (2004), also adopted by Musi et al. (2018). The quantitative analyses used Log Likelihood to test for statistically significant differences.

3. Results and Discussion

3.1. Do butCs in SUNCODAC follow the pattern described by Couper-Kuhlen and Thomson (2000) and Musi et al. (2018)?

First, we decided to check whether butCs in SUNCODAC follow the pattern (henceforth, 'canonical pattern') described by Couper-Cuhlen and Thomson (2000: 38) and Musi et al (2018). This pattern can be illustrated with example (1) taken from our corpus:

(1) You have done a great job with your translation, but I would like to make some changes. (16LV__ Alice in Wonderland 2016-A)

As can be seen in (1), Pattern A works as follows: a student first concedes the validity of another student's point (i.e. their translation proposal) in Proposition A, and they do it by means of partial agreement, approval or praise for the proposed translation A (i.e. "You have done a **great** job"). And then, in Proposition B (the denial of expectation move), this student goes on to make a potentially contrasting point by suggesting changes to the original translation. This proposition typically contains some sort of mitigated criticism or imposition (i.e. "but I **would** like to make **some** changes").

As can be seen in Table 1, our analysis revealed that 75.2% of butCs in SUNCODAC typically follow the canonical pattern. However, the analysis revealed this pattern, though prevalent in our corpus, could not account for all the instances of butCs. Thus, a corpus-based approach was adopted to detect other patterns. As a result of the manual analysis of concordance lines, two additional patterns emerged (Patterns 2 and 3), whose frequencies are also shown in Table 1.

Pattern	Frequency	%
Canonical pattern	124	75.2
Reversal of canonical pattern	19	11.5
Miscellaneous pattern	22	13.3
TOTAL	165	100

Table 1: Concessive patterns in SUNCODAC *butCs.*

In 11.5% of the instances of butCs, the order was occasionally reversed (hence the label 'reversal of canonical pattern'). This means that Proposition A is the one including the alternative translation, while Proposition B is the one containing positive evaluation, as illustrated in (2):

(2) And I chose "James R . Flynn descubrió que "instead of "reparó" but I think the verb you chose works just as well (16ASE_Intelligence_2016-B).

Finally, a miscellaneous pattern was also identified to account for variations of the preceding two patterns as in (3), where Proposition A includes the alternative translation and B is an evaluation of this alternative; or (4) where a butC appears inside another concessive headed by *although*. This heterogeneous pattern represents 13.3% of the total tokens of butCs.

(3) "Finally, it sounds better for me "largas mensulas piramidales invertidas", but maybe it is a bit stiff" (16DRP_Male_The gift of the gab_2016A).

(4) "Although this is a good translation, I would use "intentar "instead of "tartar", but it is just because it sounds more casual for me" (17AGO_The river_2017-A).

3.2. How frequently do butCs in SUNCODAC co-occur with the lexical features considered in Musi et al. (2018)?

In order to address the second research question, we explored the relative importance of the co-occurrence of butCs with first and second personal pronouns and adjectives, hedges and positive and negative words by considering the frequency of this co-occurrence. By calculating the frequency of butCs containing each of these features (Table 2).

Lexical feature	Frequency of concessives containing	% of concessives

[3] The *concessive* value of *but* has been generally ignored in the literature. For a detailed description of the *concessive*, *contrastive*, and *corrective* meanings of *but*, see Izutsu (2008).

	the feature	containing the feature
I-words	145	87.9
Positive	106	64.2
You-words	93	56.4
Hedge	85	51.5
Negative	19	11.5
We-words	15	9.1
Zero features	19	11.5

Table 2: Frequency of butCs containing at least one token of each of the selected linguistic features.

As shown in Table 2, the fact that only a small percentage of butCs contain no instances of these features seems to indicate that their presence is highly relevant in this characterization. Furthermore, the high overall incidence of butCs containing I- and you-words points to a type of discourse in which the high prevalence of first-person voice combines with the importance of the appeal to other users, as happens in texts of a dialogical nature. Also, the abundance of butCs with positive words and hedges suggests that participants in this discussion are focused on "phrasing things in such a way as to take into consideration the feelings of others" (Morand and Ocker, 2003: 2). This concern for politeness becomes particularly important in a context where the interactions typically involve assessing each other's production.

3.3. What is the distribution of these lexical features in Propositions A and B of the concessive?

Table 3 shows the relative frequency (i.e. the frequency per ten thousand words, henceforth Fpttw) of each of the above-mentioned lexical features in each of the two concessive propositions (A and B). The Log-Likelihood (also shown in the table) was used to test for the existence of statistically significant differences between the two propositions regarding the frequency of each feature.

Feature	Fpttw in A	Fpttw in B	Log likelihood
I-words	466.2	443.0	+0.18
You-words	348.8	103.2	+42.83**
Positive	345.1	121.2	+43.32**
Hedge	88.1	282.7	-20.07**
Negative	3	40.4	-0.00
We-words	25.7	30.3	-0.11

Table 3: Relative frequency of selected linguistic features in the but-corpus and in the two concessive propositions (A and B).

Our findings reveal that the typical distribution of these lexical features in SUNCODAC butCs is clearly determined by the proposition, and that this distribution directly mirrors the semantic and interactional function of each proposition. Thus, you-words and positive sentiment words feature prominently in Proposition A, while Proposition B is clearly marked by the presence of hedges. In contrast, no statistical differences were found in the case of negative words, whose low frequency may be connected with the fact that SUNCODAC participants tend to avoid overt criticism of the other participants' translations (i.e. they try to minimize threats to positive face), and also avoid presenting their alternative translations in a way that can be perceived as a threaten to their classmates' negative face (hence the occasional use of negative words to qualify their own suggestions for improvement).

4. Conclusion

As shown in the previous sections, addressing the description of concessives by referring to their role as politeness strategies is especially relevant for the study of CMC L2 contexts which include assessment and evaluation of peers' writing. We contend that, in these contexts, this characterization of concessives is relevant in terms of politeness, for three reasons: (1) Proposition B typically contains a FTA, i.e. an imposition realised as a suggestion for improvement of another student's translation; (2) this imposition is typically mitigated by means of hedging, an example of the workings of negative politeness; and (3) the FTA in Proposition B is typically preceded in Proposition A by some sort of positive politeness realised as positive evaluation or agreement with the other student. Furthermore, this characterization may afford a new insight into the use of politeness strategies not only in asynchronous online discussion forums, but also in other CMC modes as well.

Future analysis should reveal the extent to which this characterization can be extended to other types of concessives. In addition, further research will necessarily involve a refinement of the lists of lexical features which are relevant for the characterization of concession. All in all, we have tried to describe how concession and other politeness strategies work together towards "creating a comfort zone in which to exchange ideas as well as motivating students' participation" in the discussion and hence, in the learning process (Schallert et al., 2009: 715). We hope our study has contributed to a better understanding of the role of this rhetorical relation in this kind of discussion forum, but its role in other types of CMC still needs to be investigated

5. References

Barth, D. (2000). That's true, although not really, but still: Expressing concession in spoken English. In E. Couper-Kuhlen and B. Kortmann (Eds.), *Cause-Condition-Concession-Contrast: Cognitive and Discourse Perspective*. Berlin: Mouton de Gruyter, pp. 411–437.

Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Brown, P. and S. C. Levinson. (1978). Universals in language usage: Politeness phenomena. In E. N. Goody (Ed.), *Questions and Politeness*. Cambridge: Cambridge University Press, pp. 56–289.

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

Cal-Varela, M. and F. J. Fernández-Polo (2019). Preparing the ground for critical feedback in online discussions: A look at mitigation strategies. In J. Longhi and C. Marinica (Eds.), *CMC Corpora through the Prism of Digital Humanities*. Paris: L'Harmattan, pp. 15–34

Cal-Varela, M. and F. J. Fernández-Polo. 2020. SUNCODAC: A Corpus of Online Forums in Higher Education. Nexus-AEDEAN 2020/2: 44–52.

Couper-Kuhlen, E. and S. Thompson. (2000). Concessive patterns in conversation. In E. Couper-Kuhlen and B. Kortmann (Eds.), *Cause, Concession, Contrast: Cognitive and Discourse Perspectives*. Berlin: Mouton de Gruyter, pp. 381–410.

Deris, F. D., R. T. Hooi Koon and A. R. Salam (2015). Virtual Communities in an Online English Language Learning Forum. *International Education Studies* 8/12: 79–87.

Fernández-Polo, F. J. and M. Cal-Varela (2017). A Description of Asynchronous Online Discussions in Higher Education. In C. Vargas-Sierra (Ed.), *Professional and Academic Discourse: An Interdisciplinary Perspective*. Epic Series in Language and Linguistics 2: 256–264.

Fernández-Polo, F. J. and M. Cal-Varela (2018). A structural analysis of student online forum discussions. In F. J. Díaz Pérez and M. Á. Moreno Moreno (Eds.), *Looking at the Crossroads: Training, Accreditation and Context of Use*. Jaén: Universidad de Jaén, pp. 189–200

Goffman, E. (1967). *Interaction Ritual: Essays on Face-to-Face Interaction*. Chicago: Aldine.

Grote, B., N. Lenke and M. Stede. (1997). Ma(r)king concessions in English and German. *Discourse Processes* 24/1: 87–117.

Herring, S. C. (2023). Grammar and Electronic Communication. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*, pp. 1–9.

Hinkel, E. (2005). Hedging, Inflating, and Persuading. *Applied Language Learning* 15/ 1–2: 29–53

Hu, M. and B. Liu. (2004). Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177.

Hyland, K. (2005). *Metadiscourse*. London: Continuum.

Izutsu, M. N. (2008). Contrast, concessive, and corrective: Toward a comprehensive study of opposition relations. *Journal of Pragmatics* 40/4: 646–675.

Locher, M. A. (2004). Power and politeness in action: Disagreements in oral communication. New York: Mouton de Gruyter.

Locher, M. A. and R. J. Watts. (2005). Politeness theory and relational work. *Journal of Politeness Research* 1: 9–33.

Locher, M. A. and R. J. Watts. (2008). Relational work and impoliteness: Negotiating norms of linguistic behaviour. In D. Bousfield and M. A. Locher (Eds.), Impoliteness in Language: Studies on its Interplay with Power in Theory and Practice. Berlin: Mouton De Gruyter, pp. 77–99.

Mann, W. C. and S. A. Thompson. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse* 8/3: 243–281.

Morand, D. A. and R. J. Ocker. (2003). Politeness theory and computer-mediated communication: A sociolinguistic approach to analysing relational messages. *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS"03).*

Musi, E., D. Ghosh and S. Muresan. (2018). ChangeMyView Through Concessions: Do Concessions Increase Persuasion? Discourse and Dialogue 9/1:1–21.

Pyo, J. and C. H. Lee. 2019. The Effects of Mitigation Strategies Instruction in Peer Response to L2 Writing through Computer-Mediated Communication at University Level. *Multimedia-Assisted Language Learning* 22/4: 103–133.

Schallert, D., L., Y. V. Chiang, Y. Park, M. E. Jordan, H. Lee, A. J. Cheng and K. Song. (2009). Being polite while fulfilling different discourse functions in online classroom discussions. *Computers and Education* 53/3: 713–725.

Scott, Mike. 2016. *WordSmith Tools Version 7*. Stroud: Lexical Analysis Software.

SUNCODAC. (2021). *Santiago University Corpus of Discussions in Academic Contexts*. Santiago de Compostela: University of Santiago de Compostela. [http://www.suncodac.com]

Swanson, R., B. Ecker and M. A Walker. (2015). Argument mining: Extracting arguments from online dialogue. *24th Meeting of the Special Interest Group on Discourse and Dialogue* (SIGDIAL2023), pp. 217–226.

Taboada, M. and M. L. A. Gómez-González. (2012). Discourse markers and coherence relations: Comparison across markers, languages and modalities. *Linguistics and the Human Sciences* 6/1-3: 17-41.

Tanskanen, S. and J. Karhukorpi. (2008). Concessive repair and negotiation of affiliation in e-mail discourse. *Journal of Pragmatics* 40/9: 1587–1600.
Watts, R. J. 1992. Linguistic politeness and politic verbal behaviour: Reconsidering claims for universality. In R. J. Watts, S. Ide and K. Ehlich (Eds.), *Politeness in Language: Studies in its History, Theory and Practice*. Berlin: Mouton de Gruyter, pp. 43–69.
Watts, Richard. J. 2003. *Politeness.* Cambridge: Cambridge University Press.

Corpus-based didactics in higher educational settings: Empirically investigating online German youth language phenomena

Laura Herzberg, Louis Cotgrove
Leibniz-Institute for the German Language
E-mail: herzberg@ids-mannheim.de, cotgrove@ids-mannheim.de

Abstract

This paper addresses the importance of teaching digital literacy among German studies students, enrolled in teacher training, by introducing a seminar which combines Digitally-Mediated Communication (DMC) and sociolinguistics. We present real-life digital communication scenarios for research-based learning in order to increase student engagement and foster digital literacy for future teaching careers. Practical experiences in applying corpus-based methods in a German linguistics class are discussed, utilizing the NottDeuYTSch corpus and the KorAP search engine. The investigation of youth language use cases demonstrates students' ability to develop their own research projects while also putting DMC data at the forefront of authentic language use analysis.

Keywords: research-based learning, digital literacy, corpus-based methods, higher education, YouTube, youth language, DMC

1. Introduction

Students' digital literacy and ability to utilize new technologies effectively and efficiently for their learning and future careers is an obstacle in the present day, as rapid and extensive information requires individuals to adapt to a constantly changing digital world while simultaneously understanding the online cultural environment as a whole with its specific aspects of referentiality, communality, and algorithmicity (Stalder 2018: 58).

Following a questionnaire given to 30 students of German studies in Germany with a focus on teacher training, we found that students were unsatisfied with their own secondary education exposure to Digitally-Mediated Communication (DMC) and sociolinguistic topics, such as youth language. We then decided to address this issue in a seminar to provide students the necessary tools to create their own future educational materials to teach these topics to future learners.

Our research aims in this paper are as follows:

1. To present use case scenarios in everyday digital communication that can be implemented in the form of research-based learning in language classes.
2. To increase students' motivation to engage in active participation during class and scaffold their learning of advanced concepts and methods through accessible and authentic material (Wood et al. 1976; Ballis 2014; Dizon 2022).
3. To train digitally-literate students by providing hands-on materials that can be further used as a valuable resource for their own future teaching careers.

In Section 2, we explain the importance of teaching concepts against the backdrop of digital and data literacy in higher educational settings. Section 3 shows the analysis of youth language as a field of research. In the following part, Section 4, we illustrate our practical experiences from applying corpus-based methods in a class with students of German linguistics. We present our youth language use cases and describe the corpus infrastructure KorAP as well as the queried search strings the students applied on a corpus of German YouTube comments, the NottDeuYTSch corpus. Section 5 consists of concluding remarks and an outlook.

2. Corpus-based methods in language teaching: Digital and corpus literacy

German linguistics has a well-developed landscape of corpora at its disposal for the systematic and usage-based study of the German language, e.g. DeReKo,[1] the DWDS Corpora, [2] and DeTenTen. [3] These corpora are made available on a sustainable basis via digital resource infrastructures and can be researched and queried via web-based user interfaces. These offer a wide range of possibilities for investigating levels of analysis and contexts of language use relevant to schools: Lexicon, semantics, word formation, or syntax.

"Digital literacy is the ability to properly use and evaluate digital resources, tools and services, and apply it to lifelong learning processes" (Gilster 1997: 220).

We understand corpus literacy, viewed as subcategories of digital literacy, as a competence skill set which can be applied when studying a language. From this perspective, corpora, even if they are primarily provided as resources for scientific research, come into play as educational technologies. Through their systematic implementation in teaching contexts, they open up a multitude of possibilities for the study of language.

DMC is not only a subject to be talked about in class but it used an empirical data base to analyse authentic language use. The aim is to familiarize students with the testing of hypotheses about the German language as a

[1] https://www.ids-mannheim.de/digspra/kl/projekte/korpora

[2] https://www.dwds.de/r

[3] https://www.sketchengine.eu/detenten-german-corpus/

process based on the systematic evaluation of data and to develop an understanding of linguistics as an empirical field of science. Additionally, students get to know and abstract a familiar area of their everyday life from a scientific perspective. They understand the broader differences between oral and written language use, conversations, texts and DMC.

3. Youth language as a field of research

Similar to DMC, youth language can function as an engaging framework for many students of language to reflect on language use in society and develop their scientific skills and methods within the context of higher education. Language plays a pivotal role in shaping youth identities during the transitional stage of adolescence (Stenström et al. 2002; Holmes 2013). It serves as a symbolic assertion of autonomy, allowing young people to affiliate with or distance themselves from relevant peer groups and youth subcultures (Bahlo 2019). Linguistic innovation emerges as a departure from mainstream norms and dominant cultural values, reflecting the friction between the worlds of 'youth' and 'adulthood' (Androutsopoulos and Georgakopoulou 2003; Dürscheid, Wagner, & Brommer 2010; Tagliamonte 2016). Adolescents use language as a means of expressing their individuality, forging connections with like-minded peers, and carving out their unique identities within the broader societal landscape.

The fields of youth language and DMC are also strongly intertwined, due to young people's experiences with digital communication from an early age (Bahlo et al. 2019: 80). However, researchers must be careful not to conflate features of DMC with features of youth language: not all non-standard language is produced as part of performing youth identity; sometimes it is a characteristic of the online medium, which happens to be particularly frequently used by young people (cf. Dürscheid and Spitzmüller 2006). Secondly, youth language on DMC varies based on platform: the findings of previous research on SMS or instant messaging clients (e.g. Nowotny 2005; Kleinberger Günther/Spiegel 2006), are not always applicable to studies of Facebook (back when it was popular among young people, e.g. Androutsopoulos 2015) or WhatsApp (e.g. Siebenhaar 2018). This highlights the need for more nuanced investigations into the relationship between language use, digital media, and age demographics, but it also means that the subject is of particular interest to students, who can feel closer to the source material than their older tutors.

4. Practical experiences from a university seminar

This section outlines the use case we selected to develop sociolinguistic awareness and methodological competence of corpus linguistics in our students, as well as a more in-depth description of the appropriateness of the corpus for our research aims.

We decided to initially focus on discursive differences between similar linguistic patterns in DMC youth language, as this offered students the opportunity to learn the functionalities of our chosen corpus and corpus tool (in this case querying the NottDeuYTSch corpus with KorAP, explained in Sections 4.1 and 4.2), whilst being able to apply their own knowledge as (mostly) experts of the sociolect to the situation, and are interesting interdisciplinary fields of research for students.

The selected discursive topic enabled the investigation of differences between self-description, the communication with other commenters, parasocial communication with YouTuber personalities, as well as attribution of traits to 3rd parties.

4.1 Data base: The NottDeuYTSch corpus in KorAP

Within the scholarship on the linguistic aspects of digital media communication), there has been a prevalent assumption that the language use under study is that of young people. However, this assumption is often not explicitly stated (Androutsopoulos 2003a; Siebenhaar 2018). This is partly due to the ethnographically sparse nature of much online data, making it challenging for researchers to conclusively establish the demographics of their data sources. However, YouTube comments represent an untapped source of authentic language data from young people. Despite YouTube being one of the most accessed means of communication for this demographic for several years (Saferinternet.at 2018; Bahlo et al. 2019), studies on digital media communication (DMC) or youth language have rarely analyzed the linguistic features used by young people in YouTube comments. This is surprising, given that 77% of German 14-19 year-olds describe themselves as active users of YouTube (Statista 2023a). The *NottDeuYTSch* corpus provides this data, containing over 33m words written by young people in YouTube comments under mainstream German-language videos aimed at a young audience between 2008 and 2018.

Statistic	**Value**
Number of Tokens (including emoji and emoticons)	33,760,494
Number of Tokens (only lexemes)	32,549,462
Number of Types	567,086
Type-Token Ratio (TTR)	0.017
Number of Comments	3,149,457
Number of Videos	296
YouTube Channels Represented	63
Mean Tokens per Comment	10.720
Median Tokens per Comment	5
Mean Comments per Video	1,914

Table 1: Statistical overview of the NottDeuYTSch corpus

Using the NottDeuYTSch corpus can improve motivation and Willingness to Communicate (WTC) in student learners, due to the high levels of exposure and familiarity they have to YouTube and DMC environments. The corpus can also be used to develop a pragmatic and

metacommunicative understanding of DMC texts and interactions, and the size and structure of the corpus enables microdiachronic and genre analyses, and is suitable for a wide variety of linguistic research that can be carried out by students, whether lexical, syntactical, morphological, metalinguistic, multilingual, or discursive.

The search engine KorAP (*corpus analysis platform*) (Bański et al. 2013; Diewald et al. 2016) has been developed as the main access point to DeReKo, the German Reference Corpus.

The NottDeuYTSch corpus can be accessed via the KorAP search engine after registering.

4.2 Practical application of youth language use cases

In an empirical linguistic teaching environment, students of German linguistics (N=32) in their advanced semesters analysed comment sections of youth-oriented YouTube videos by applying corpus-based methods, querying a variety of search strings on the NottDeuYTSch corpus via the KorAP search engine.

A total of 32 students took part in the seminar. They were in their 4th to 8th semester, i.e. in the second half of their Bachelor's degree course. At this point, they had already attended basic introductory courses and were able to choose from various in-depth specialist seminars in German linguistics. The students come from three-degree programs: Bachelor of German Studies: Language, Literature, Media; Bachelor of Education Teaching Qualification "Gymnasium" (*secondary school*) and Bachelor of Culture and Business: German Studies. The majority of the students had basic knowledge in the field of corpus linguistics, i. e. they are familiar with the concepts of corpus linguistic research, but have not actively used such systems themselves. None of them had worked with the search engine KorAP beforehand.

Our chosen use case analysed the typical DMC youth language construction [X SEIN so ART X] (X BE such ARTICLE X)[4], as can be seen in Example 1. This construction is found over 10 times more frequently in DMC youth language than in comparable DMC corpora over the same time frame and 20 times more frequently than standard written language.[5]

(1) bibi, du bist so eine tolle Person
[bibi, you are such an amazing person]
NDY/258/010904

After registration and solving entry-level queries in order to get to know the search engine, the students queried the following German search strings in KorAP:

1. du bist so ein
[you are such a.MASC/NEUT]
2. [tt/l=sein] so ein
[lemma=be | so | a.MASC/NEUT]
3. [tt/l=sein] [orth=so] [tt/p=ART] [tt/p=NN | tt/p=NE]
[lemma=be | so | Part of Speech (PoS)=article | PoS=noun]
4. [marmot/m=person:1 & tt/l=sein] [orth=so] [tt/p=ART] [tt/p=NN | tt/p=NE]
[lemma=be & person=1st | so | PoS=article | PoS=noun]
5. [marmot/m=person:2 & tt/l=sein] [orth=so] [tt/p=ART] [tt/p=NN | tt/p=NE]
[lemma=be & person=2nd| so | PoS=article | PoS=noun]
6. [marmot/m=person:3 & tt/l=sein] [orth=so] [tt/p=ART] [tt/p=NN | tt/p=NE]
[lemma=be & person=3rd | so | PoS=article | PoS=noun]
7. [tt/l=sein] [orth=so] [tt/p=ART] []{0,2} [tt/p=NN | tt/p=NE]
[lemma=be | so | PoS=article | X (0-2 words) | PoS=noun]

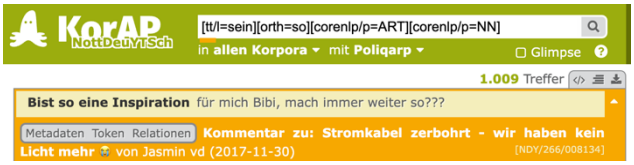


Figure 1: Query 1 in KorAP with example[6]

Simple queries such as 1., "you are", yielded interesting differences in the attribution of character traits, cf. Figure 1: The query "you are so" resulted in a wide spectrum of superlatives, from very positively connotated uses, as in Example (1), to a usage in more negative contexts, cf. Example (2).

(2) Du bist so ein Lügner
[You are such a liar]
NDY/296/007124

(3) Also Dagi !❤️ Du bist so eine starke Frau ! Du hast so viel durchgemacht!
[So Dagi !❤️ You are such a strong woman ! You have been through so been through so much!]
NDY/264/005817

Queries 2-7 then progressively expanded students' knowledge of parts of speech and verb conjugation and how one can build a query to compare the sentiment or discursive differences between verb person, as well as accounting for potential future comparative studies based on adjectival use, as demonstrated in Query 7 in Example (3). Using the queries of authentic language, the students can start to develop and test their own hypotheses on youth themselves with the search engine.

[4] There are other specific features of youth language, such as the use of intensifiers, quotatives, swear words, vague language, foreign terms, invariant tags or small clauses. We focus in this paper on "X BE such ARTICLE X" as this is an example on how students can gradually approach more difficult queries, cf. queries (1-7) while familiarizing

[5] NottDeuYTSch corpus: 130 instances per million words (ipm). DWDS WebXL corpus: 11.1 ipm. DWDS Contemporary corpus: 6.3 ipm.

[6] Translation: "You are such an inspiration to me Bibi, keep up the good work".

language beyond simple lexical queries. KorAP also enables the results to be exported as a CSV file for further investigation, for example (micro-)diachronic study of the results.

In addition to in-class feedback, we developed a questionnaire following models developed for gauging student motivation in the classroom (see Wain et al. 2019; Lee and Lu 2021), asking for students' self-evaluation of their development and understanding of corpus linguistic methods and the investigation of youth language and DMC, to refine our pedagogical methods for coming semesters.[7]

5. Conclusion and outlook

In conclusion, our investigation into students' digital literacy and their exposure to digitally-mediated communication (DMC) and sociolinguistic topics underscores the critical need for educational initiatives that equip learners with the necessary skills to navigate the complexities of the digital landscape. Through our seminar, we identified shortcomings in students' secondary education experiences and endeavored to address these gaps by empowering them by firstly, getting to know corpus-based methods and secondly, creating educational materials for their future careers.

Our research objectives centered on three main pillars: presenting real-life use cases for digital communication in language learning, fostering student motivation and active participation through authentic materials, and cultivating digitally-literate educators through hands-on training. By integrating corpus-based methods and youth language research into our curriculum, we aimed to enhance students' sociolinguistic awareness and methodological competence. Our seminar experience exemplified the efficacy of integrating corpus linguistics and sociolinguistic research into language education. By engaging students in hands-on analysis of youth-oriented digital discourse, we not only enhanced their linguistic competencies but also nurtured their curiosity and analytical skills. We understand that corpora can quickly go out of date, particularly for a dynamic sociolect that experiences rapid lexical and linguistic change, such as youth language.

However, the methods presented here can be easily adapted to more recent or other appropriate corpora or expanded to incorporate other linguistic patterns and queries, such as intensifiers and quotatives. Moving forward, initiatives like ours are essential for equipping students with the digital literacy skills necessary for success in an increasingly interconnected world.[8]

6. References

Androutsopoulos, Jannis. 2015. 'Networked Multilingualism: Some Language Practices on Facebook and Their Implications'. *International Journal of Bilingualism* 19 (2): 185–205. https://doi.org/10.1177/1367006913489198.

Androutsopoulos, Jannis, and Alexandra Georgakopoulou. 2003. 'Discourse Constructions of Youth Identities: Introduction'. In *Discourse Constructions of Youth Identities*, edited by Jannis Androutsopoulos and Alexandra Georgakopoulou, 110:1–26. Philadelphia: John Benjamins.

Bahlo, Nils, Tabea Becker, Zeynep Kalkavan-Aydın, Netaya Lotze, Konstanze Marx, Christian Schwarz, and Yazgül Şimşek. 2019. *Jugendsprache: Eine Einführung*. Berlin: J.B. Metzler.

Ballis, Anja. "Puschkin oder Podolski? - Schreiben in der Zweitsprache." In *Zweitspracherwerb im Jugendalter*, edited by Bernt Ahrenholz and Patrick Grommes, 211–30. Berlin: De Gruyter, 2014.

Bański, Piotr, Joachim Bingel, Nils Diewald, Elena Frick, Michael Hanl, Marc Kupietz, Piotr Pęzik, Carsten Schnober and Andreas Witt. 2013. KorAP: The new corpus analysis platform at IDS Mannheim. In Zygmunt Vetulani & Hans Uszkoreit (eds.), Proceedings of the 6th Conference on Language and Technology (LTC-2013). Poznań: Uniwersytet im. Adama Mickiewicza w Poznaniu. https://nbn-resolving.org/urn:nbn:de:bsz:mh39-32617.

Diewald, Nils, Michael Hanl, Eliza Margaretha, Joachim Bingel, Marc Kupietz, Piotr Bański and Andreas Witt. 2016. KorAP architecture: Diving in the deep sea of corpus data. International Conference on Language Resources and Evaluation (LREC) 10, 3586–3591.

Dizon, Gilbert. "YouTube for Second Language Learning: What Does the Research Tell Us?" *Australian Journal of Applied Linguistics* 5, no. 1 (April 30, 2022): 19–26. https://doi.org/10.29140/ajal.v5n1.636.

Dürscheid, Christa, and Jürgen Spitzmüller. 2006. *Perspektiven der Jugendsprachforschung Trends and Developments in Youth Language Research*. Sprache, Kommunikation, Kultur, Bd 3. Frankfurt am Main: Peter Lang.

Dürscheid, Christa, Franc Wagner, and Sarah Brommer. 2010. *Wie Jugendliche schreiben: Schreibkompetenz und neue Medien*. Walter de Gruyter.

Gilster, Paul. 1997. Digital literacy. New York: Wiley Computer Pub.

Kleinberger Günther, Ulla, and Carmen Spiegel. 2006. 'Jugendliche Schreiben Im Internet: Grammatische Und Orthographische Phänomene in Normungebundenen Kontexten'. In *Perspektiven Der Jugendsprachforschung- Trends and Developments in Youth Language Research*, edited by Christa Dürscheid and Jürgen Spitzmüller, 101–15. Frankfurt am Main: Peter Lang.

Lee, J. S., & Lu, Y. 2021. L2 motivational self system and willingness to communicate in the classroom and extramural digital contexts. *Computer Assisted Language Learning*, *36*(1–2), 126–148. https://doi.org/10.1080/09588221.2021.1901746.

[7] As of writing, the authors are still awaiting the completion of the questionnaires by the students.

[8] This is an ongoing investigation and we are looking forward to presenting our research and extensive evaluation of the participating students at the conference.

Nowotny, Andrea. 2005. 'Daumenbotschaften. Die Bedeutung von Handy Und SMS Für Jugendliche'. *NetWorx*, no. 44. http://www.mediensprache.net/networx/networx-44.pdf.

Siebenhaar, Beat. 2018. 'Funktionen von Emojis und Altersabhängigkeit ihres Gebrauchs in der WhatsApp-Kommunikation'. In *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, edited by Arne Ziegler, 749–72. Berlin: De Gruyter.

Stalder, Felix. 2018. The digital condition. Cambridge: Polity Press.

Tagliamonte, Sali A. 2016. *Teen Talk: The Language of Adolescents*. Cambridge: Cambridge University Press.

Wain, J., Timpe-Laughlin, V., & Oh, S. (2019). Pedagogic Principles in Digital Pragmatics Learning Materials: Learner Experiences and Perceptions. *ETS Research Report Series*, *2019*(1), 1–21. https://doi.org/10.1002/ets2.12270

Wood, David, Jerome S. Bruner, and Gail Ross. "The Role of Tutoring in Problem Solving." *Journal of Child Psychology and Psychiatry* 17, no. 2 (April 1976): 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.

Testing the weak-tie hypothesis with social media

Mikko Laitinen, Masoud Fatemi
University of Eastern Finland
E-mail: mikko.laitinen@uef.fi, masoud.fatemi@uef.fi

Abstract

This article combines the study of large-scale social media data with social network theory in sociolinguistics. Given that the purpose of social media is to form networks and communities, big data from social media applications could have substantial potential in deepening the understanding of the role of networks in language variation and change. The study first presents an algorithmic method suitable for directed-graph ego networks in computer-mediated communication. This method measures network strength and enables us to enrich social media data with a network parameter that indexes how strongly (or loosely) people in a network are connected to each other. We then use a large dataset of c. 4.8 billion words from nearly four thousand networks to study how network strength conditions linguistic change. The results show that online networks are highly similar to traditional offline networks, a finding that enables fixing a major methodological limitation in the study of weak ties, namely that the method is less suited for studying socially and geographically mobile individuals. This finding makes it possible to apply the theory of social networks in sociolinguistics to very large digital networks in social media.
Keywords: social media, network theory, ego networks, ongoing change

1. Introduction

This article presents a study that uses large-scale social media data to test weak-tie hypothesis in sociolinguistics. The hypothesis, based primarily on small-scale data from ethnographic observations, has been influential in the study of language variation and change (Milroy, 1987). It predicts that individuals form communities of varying strength, which influence the extent to which members can access novel information. Weak-tie environments have been observed to be open to external influences, facilitating change, but networks that are characterized by strong ties lead to norm-enforcing communities that resist change.

In sociolinguistics, the hypothesis has not been tested using social media data, despite a clear need (Georkakopoulou, 2011; Laitinen et al., 2020; Zhu & Jürgens 2021; Würschinger, 2021). The reason is that the methodological toolbox limits the study of offline networks to circa 30–50 individuals (Milroy & Milroy, 1992: 5) and the methods are less suited for tracing socially and geographically mobile individuals. This methodological constraint is potentially serious, because it limits the study to cover only a small portion of networks. Prior studies in social anthropology have shown that average human networks are substantially larger, with an average size well over 100 people (McCarty et al., 2001; Dunbar, 2020).

This article has a dual function. On the one hand, it extends the study of social media data in sociolinguistics to social networks. Secondly, it presents a novel algorithmic method used to enrich large-scale social media data with an interactional parameter to enable large-scale quantitative study of digital networks. Given that the purpose of social media is to form communities, the study potentially opens up new horizons in using network evidence from social media in the study of language variation and change.

The study aims at answering two research questions:

(1) Does the weak-tie hypothesis hold in large-scale data from social media?
(2) Can we improve the predictive power of the weak-tie hypothesis using quantitative evidence from social media?

The structure of the article is such that Section 2 introduces the theoretical framework, while Section 3 presents the material and methods. The main findings are visualized in Section 4, which is followed by the discussion and some thoughts on future research prospects.

2. Theoretical framework

Our work is situated within social network theory in sociolinguistics (Milroy & Llamas, 2013). The basic concept is an ego network, which forms around an anchor individual (Fig. 1). The concept is a practical tool that enables an analysis to be rooted to a specific anchor. One beneficial feature in digital applications is that individuals come with an identifier (a user ID). Other individuals connected to the ego are alters (A–E). We use data from one application that enables us to observe who is connected to whom and who communicates with whom. Our network data therefore form a directed-graph network, and the arrowed lines in Figure 1 show the interactions between the ego and alters (first-order contacts), and those between the alters (second-order ties).

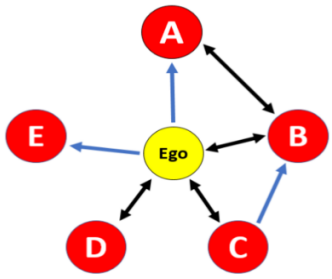


Figure 1: A directed graph network with unidirectional (blue) and bi-directional (black) edges

Networks can vary in terms of size and structure, and a network that consists chiefly of strong and multiplex ties tends to inhibit change and support norms. This is because people in close-knit surroundings do not want to "rock the boat" and risk their social standing. If connections are loose and uniplex, the members have a greater likelihood of having a more insignificant ties (acquaintances instead of friends), then innovations and influences can flow more freely (Milroy & Milroy, 1992).

To establish network strength scores, ethnographic approaches rely on surveys and interviews, but given that our data come from social media, we rely on directly observable interaction parameters. Observing authentic behavior online means that we use data from relationships that people actually have, rather than relying on impressionistic knowledge of remembering connections. The next section introduces our algorithmic approach.

3. Material and methods

The material was collected as part of a project funded by the Research Council of Finland in early 2023. The objective was to create four massive datasets that contain both texts keyed in by users and the interactional metadata to be subjected to network analysis. We used Twitter Academic API, which was closed in May 2023. The data collection targeted genuine human accounts, and by imposing a number of filters, it excluded celebrities, business, and organizations (Laitinen et al., 2020).

The four datasets represent different geographic areas (Australia, the UK, US, and the Nordic region). Twitter provides two types of location-sharing options in addition to any location details that users might include in their tweets. The first is a free text field in the user's profile where they can enter any location, real or imagined. The second option is a location tag that appears on tweets if the user enables this feature in their profile settings. This tag, provided by Twitter, includes standardized coordinates in terms of latitude and longitude.

3.1 Data collection

The final dataset was collected through several phases (see Fig. 2). The starting point was to set up data streams in the four areas, which was followed by extracting user information, cleaning and filtering the data, and lastly collecting the messages and the networks of each user. Essentially, we compiled an initial list of potential users from the four main geographical areas (AU, UK, US, and Nordic), refined the lists through various filters, and collected networks and tweets for the accounts that remained.

During the filtering stage, we first set a threshold to remove accounts that were either very inactive or overly active, specifically excluding users who tweeted less than three or more than 12,000 messages per year on average. Second, we removed celebrities and policymakers, typically verified users with millions of followers but few friends. Third, to improve the accuracy of location tags, we included only those users whose self-reported profile locations matched the locations indicated in their tweets. Lastly, we eliminated accounts with more than 500 contacts (friends and followers combined) from our list of potential candidates. This last step aims at ensuring that we investigate average human networks (Dunbar 2020).

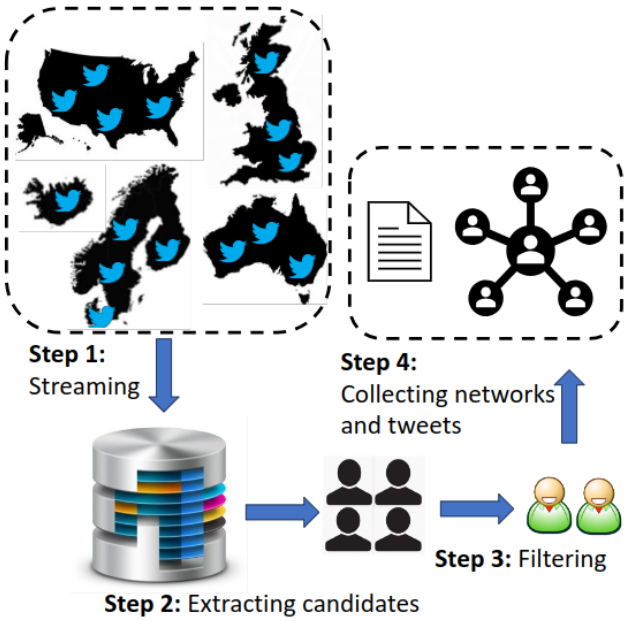


Figure 2: data collection process

After completing the filtering process, we collected all the user-generated material (up to 3200 recent messages) from all the users (both ego nodes and their alters) involved in each network. The dataset includes 19,345 ego networks involving 759,495 individuals and nearly 11 billion words from social media posts. Since this article serves as a proof-of-concept, this study is limited to data collected from the US and UK. It consists of about 4.8 billion words from 233,774 users across 3,935 networks (Table 1). These are divided into urban inner city and suburban/rural areas based on their location, as detailed in Table 1.

The dataset includes nearly 1.8 million connections between all the accounts in each of the 3,935 ego networks. This interactional information of who is a friend with whom, whose messages are shared, replied, or quoted, and how many times an account interacts with the others.

Table 1: The basic frequencies in the datasets

	Egos	Alters	Tokens	Mean size
US urban	2,037	123,031	2,445,454,135	61
US sub/rural	958	56,453	1,163,547,607	59
UK urban	538	30,168	619,904,835	57
UK sub/rural	402	24,122	554,319,126	60
Total	3,935	233,774	4,783,225,703	60

Table 1 also shows that the mean network size of our data is 60, so well over the methodological upper limit of 50 people in traditional network analysis.

3.2 Extracting network information

The data were subjected to algorithmic processing to calculate a network strength label for each network (Laitinen et al., 2020; Laitinen & Fatemi, 2022). It consists of a multidimensional process that takes into account six measures extracted from each ego network in the dataset.

3.2.1 Interaction strength

Interaction strength (IS) captures the interactions between each pair of nodes in the network that are connected via an edge. On Twitter, interactions between users occur through likes, replies (mentions), retweets, quotes (retweeting with added text), and direct messages. However, due to privacy constraints, accessing the lists of likes or direct messages using the Twitter API is impossible. For a directed edge from node *a* to node *b*, the weight of the edge is calculated as follows:

$$W_{ab} = (w_1.retweet) + (w_2.quote) + (w_3.reply)$$
$$\text{where } \sum w_i = 1 \qquad (1)$$

Here, the terms retweet, quote, and reply represent the number of interactions from node *a* to node *b* through retweeting, quoting, or replying to node *b*'s tweets, respectively. The values w_1, w_2, and w_3 are regulatory weights that adjust the influence of each type of interaction. Once we have calculated the IS values for all the edges in the network, we determine the average IS to represent the final value for the overall network.

3.2.2 Average and range of betweenness centrality

For the subsequent three measures, we implemented centrality concepts. In graph theory and network analysis, centrality measures calculate the influence or centrality of a node or edge within a network. We focused on two specific measures: betweenness and closeness centrality, from which we extracted three distinct features for our analysis. Betweenness centrality (BC) calculates to what extent a node falls on the shortest path between every other pair of nodes in the network. For a node like *a,* the BC calculates as follows:

$$BC(a) = \frac{1}{(N-1)(N-2)} \sum_{u,v \neq a \in V} \frac{\sigma(u,v|a|)}{\sigma(u,v)}$$

Here *N* represents the total number of nodes in the network, $\sigma(u,v|a)$ indicates the number of shortest paths between nodes *u* and *v* that pass through node *a*, and $\sigma(u,v)$ denotes the total number of shortest paths between nodes *u* and *v*. After computing BC values for every node, we first calculate the average betweenness centrality (ABC) value across the entire network. Subsequently, we determine the range of betweenness centrality (RBC) values, which is the difference between the highest and lowest BC values in the network.

3.2.3 Asymmetric closeness centrality

Closeness centrality measures how close a node is to every other node in a network by using the shortest paths to calculate this proximity. Closeness centrality values range between 0 and 1, with a higher value suggesting that a node is closer to the rest of the network (i.e., can reach them more quickly). In a directed network, each node, such as node *a*, has two closeness centrality values, incoming and outgoing, calculated as follows:

$$CC_{in}(a) = \frac{N-1}{\sum_{a \neq b \in V} dist(b,a)} \qquad (3)$$

$$CC_{out}(a) = \frac{N-1}{\sum_{a \neq b \in V} dist(a,b)} \qquad (4)$$

Here, *dist* is the function that calculates the distance between two nodes in a network by counting the number of edges in the shortest path connecting them. After determining the incoming and outgoing closeness centrality values for all nodes in the network, we compute two average values: one for average incoming CC and the other for average outgoing CC. The difference between these two average values gives us the final asymmetric closeness centrality (ACC) value for the network.

3.2.4 Social similarity

For this measure, we applied the Jaccard Similarity principle, creating a feature that captures the number of mutual friends between each pair of nodes in the network. Precisely, the social similarity (SS) for a the entire network is calculated as follows:

$$SS = \frac{2}{N(N-1)} \sum_{a,b \in V} \frac{\left|a_{friends} \cap b_{friends}\right|}{\left|a_{friends} \cup b_{friends}\right|} \quad (5)$$

After computing the SS values for all nodes in the network, we then determine the average SS value to represent the overall network.

3.2.5 Outliers

The last measure we extracted is called outliers (OUT). It calculates the percentage of nodes in an ego network that become isolated when the ego node and its connecting edges are removed. Higher values indicate that a larger portion of the nodes are connected solely through the ego node, suggesting a scenario of weaker ties within the entire ego network.

We extracted these six measures (IS, ABC, RBC, ACC, SS, and OUT) from each ego network in our dataset and then normalized each feature using min-max normalization. For features like IS and SS, higher values indicate a stronger tie network. Conversely, measures such as OUT, ACC, ABC, and RBC represent a stronger tie network with lower values. Therefore, we used reverse scaling on these to align higher values with stronger ties. Finally, we calculated the average of these six adjusted features for each ego network to determine the network score (NS).

The process results in NS that can theoretically range between 0 and 1. If the mean is 0 (practically impossible), it indicates a maximally loose-tie network, while the value of 1 indexes a maximally close-knit network, where every node is connected to each other.

As a result, each one of the 3,935 ego networks are labelled with a figure that indexes how strongly- or loosely-connected the nodes in a network are.

3.3 Programming stack

We primarily used Python programming language for data collection, preprocessing, feature extraction, and calculating network scores. For statistical analysis and feature extraction, we employed Python libraries such as Pandas and NumPy. Additionally, we used Networkx for network analysis and creating visualizations, and Spacy for adding part-of-speech tags to our data.

4. Results

Our case study deals with whether online networks are similar to offline networks, which have been studied in previous studies in sociolinguistics. We test if we can observe digital weak-tie environments to have a greater frequency of the incoming linguistic forms than close-knit networks. If the answer is positive and we can confirm the weak-tie hypothesis, it indicates that computer-mediated networks at least on this social media application are similar to offline environments.

We operationalize the study in two ways. First, through variables that can contain two or more ways of saying the same thing (Tagliamonte, 2012). Variables are not categorical, but investigating quantitative tendencies can reveal meaningful patterns. The linguistic variable used here is contacted forms of verbs (e.g. are > 're, etc.) and negators (e.g. not > n't). It has been selected because contractions are undergoing change in contemporary English. Mair (2006: 189) suggests that the shift towards contracted forms is "strongest in American English". The change has been linked with colloquialization, the gradual shift of spoken norms into written language (Leech et al. 2009). We should be expected to observe frequent use of contractions in computer-mediated communication, given that social media communication is highly interactional (Knight et al 2014). Any pattern in which background parameters, such as network strength, correlates with the frequency of use, is likely to be brought about and conditioned by the background variable. Second, we use the normalized frequencies of English semi-modal NEED *to* V-inf (e.g. You need to do this). It is also undergoing change, and is "spectacularly increasing" in recent English (Leech et al., 2009: 94; Mair, 2015).

Given that we are interested in large-scale patterns, the forms are automatically retrievable in parts-of-speech-tagged material and frequent enough for analysis. Lastly, they concern distinct linguistic categories, contractions of orthography and the semi-modals of grammar, meaning that they differ in terms of complexity in processing. Contractions are most likely above the level of linguistic awareness and might also be corrected or suggested by proof-reading and text-generating devices today, while NEED *to* + V-inf is more complex, and less likely to be known or directly analyzable by language users in fast-paced digital communication on social media.

The results show two things. On the one hand, they confirm that small online networks are highly similar to offline environments that have been studied in prior sociolinguistic studies. As shown in Table 1 above, the mean size of network in our dataset is around 60 nodes, and we use that as an approximate threshold, so any network below that is considered to be small, and vice versa. Figure 3 illustrates a key finding with contractions in all the 1,704 small networks. It shows that people in weak-tie environments use more contracted forms as could be expected according to the hypothesis. This observation holds even in computer mediated communication. We have only included the observations obtained from contractions, but the same broad pattern also holds for the greater normalized frequencies of NEED *to* V-inf. Weak-tie environments systematically show greater frequencies of the incoming forms.

Figure 3: Contractions in 1,704 small US networks

On the other hand, the results also suggest that when we start to increase the network size to larger networks (>60 members and above), the difference between weak- and strong-tie environments starts to disappear. Here, it is important to keep in mind that prior sociolinguistic approaches have limited the analysis to <50 member networks. A key finding are shown in Figure 4 below. Here, we use a conservative threshold of 60 members are the cut-off point, but the finding can also be replicated in networks of 50 people or more.

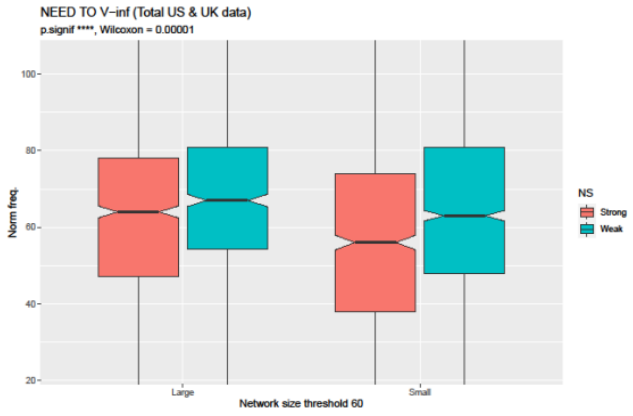


Figure 4: NEED *to* V-inf in the entire data

Figure 4 shows that the differences between weak- and strong-tie environments start to disappear, so that in small networks, language users in weak-tie environments use the incoming linguistic item in greater frequencies than those in close-knit surroundings (the boxplots of the right). The difference is not as pronounced in large networks (the boxplots on the left). The differences are not as pronounced when we observe the data from contractions (Fig 5 showing the results from all the data).

The results presented here are limited to two linguistic structures, but they are potentially highly significant. The fact that we can replicate the key finding of the weak-tie hypothesis in extremely large digital datasets only strengthens the hypothesis. The findings suggest that data from this social media application are surprisingly similar and behave similarly to observational data from other settings that have been investigated in previous sociolinguistic studies. Various studies have suggested that the size and structure on online networks closely resemble offline settings (Gonçalves et al. 2011; MacCarron, Kaski, Dunbar 2016). The observations presented in our study suggest that the same applies to networks as conditioning elements in innovation diffusion.

Given that social media connects large numbers of people networks from a range of settings, and the fact that it could provide a test bed for the weak-tie hypothesis is substantial. Social media data with large scale and scope could be used as a testing ground for the network theory.

5. Discussion and future research outlook

This study extends the study of computer mediated communication to social networks in sociolinguistics. It first of all presents an algorithmic methodology to measure network strength in digital networks. As said, not all social media applications result in directed graph networks, but the study has potential for those that do so. It is clear that we can access a far greater variety of networks using social media data and data-intensive than using manual methods for data collection. Second, the results show that evidence from social media applications and computer mediated

communication can substantially contribute to old discussions of the role of social networks in language variation and change.

The results clearly suggest that online networks are highly similar to offline networks when it comes to them conditioning linguistic change. That is, small weak-tie networks show systematically higher frequencies of incoming linguistic features than strong-tie networks. Equally important is the fact that we see evidence of this difference disappear when network size grows large. This finding suggests that large-scale social media data have substantial potential in studying social networks.

Various studies have focused on internet neologisms in network studies (e.g. Del Tredici & Fernández 2018; Zhu & Jürgens 2021), but we have shown that the weak-tie hypothesis also holds for variables that are advanced, but clearly ongoing diachronic change. It is clear that the weak-tie hypothesis not only refers to incipient change and innovation, but is also applicable to broader macro-level change (Milroy & Milroy 1985).

6. Copyrights

All data presented here have been obtained using the Academic API in 2016–2023. They are available for fundamental research under the European Union's Data Directive (EU) 2019/790 on copyright in the Digital Single Market. All visualizations are ours.

7. References

Dunbar, R. (2020). Structure and function in human and primate social networks: Implications for diffusion, network stability and health. *Proceedings of Royal Society A*. London. 476A. doi: 10.1098/rspa.2020.0446.

Georgakopoulou, A. (2011). Computer-mediated communication. In J-O. Östman & J. Verschueren (Eds.), *Pragmatics in Practice*, 93–110. Amsterdam: John Benjamins.

Gonçalves, B., Perra, N. & Vespignani, A. (2011). Modeling users' activity on Twitter networks: Validation of Dunbar's Number. *PLoS ONE*, 6:8: e22656. doi: 10.1371/journal.pone.0022656

Knight, D., Adolphs, S. & R. Carter. CANELC: Constructing an e-language corpus. *Corpora* 9(1), pp. 29–56.

Laitinen, M., Fatemi, M. & Lundberg, J. (2020). Size matters: Digital social networks and language change. *Frontiers in Artificial Intelligence* 3:46. doi: 10.3389/frai.

Laitinen, M. & Fatemi, M. (2022). Big and rich social networks in computational sociolinguistics. In P. Rautionaho, H. Parviainen, M. Kaunisto & A. Nurmi (Eds.), *Social and Regional Variation in World Englishes: Local and Global Perspectives*. London: Routledge, pp 166–189. doi: 10.4324/9781003227342-9.

Laitinen, M., Lundberg, J., Levin, M., & Martins, R. M. (2018). The Nordic Tweet Stream : A Dynamic Real-Time Monitor Corpus of Big and Rich Language Data. DHN 2018 Digital Humanities in the Nordic Countries 3rd Conference : Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference Helsinki, Finland, March 7-9, 2018, 349–362.

Leech, G., Hundt, M. Mair, C. & Smith, N. (2009). *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.

MacCarron, P., Kaski, K. & Dunbar, R. (2016). Calling Dunbar's numbers. *Social Networks*, 47, pp. 151–155, doi: 10.1016/j.socnet.2016.06.003

Mair, C. (2015). Cross-variety diachronic drifts and ephemeral regional contrasts: An analysis of modality in the extended Brown family of corpora and what it can tell us about the New Englishes. In P. Collins (Ed.), Grammatical Change in English World-wide. Amsterdam: John Benjamins, pp. 119–146.

Mair, C. (2006). *Twentieth-Century English: History, Variation, and Standardization*. Cambridge: Cambridge University Press.

McCarty, C., Killworth, P. D., Bernard, H. R., Johnsen, E. C. & Shelley, G. A. (2001). Comparing two methods for estimating network size. *Human Organization* 60(1), pp. 28–39.

Milroy, L. (1987). *Language Change and Social Networks*. 2nd edition. Oxford: Blackwell.

Milroy, L. & Milroy, J. (1992). Social network and social class: Toward an integrated sociolinguistic model. *Language in Society* 21, pp. 1–26.

Milroy, J. & Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics*, 21, pp. 339–384.

Milroy, L. & Llamas C. (2013). Social networks. In J.K. Chambers & N. Schilling (Eds.), *The Handbook of Language Variation and Change*. Oxford: Blackwell, pp. 409 – 427.

Tagliamonte. S. (2012). *Variationist Sociolinguistics. Change, Observation, Interpretation*. Oxford: Blackwell.

Del Tredici; M. & Fernández, R. (2018). The road to success: Assessing the fate of linguistic innovations in online communities. In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 1591–1603, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Würschinger, Q. (2021). Social networks of lexical innovations. Investigating the social dynamics of diffusion of neologisms on Twitter. *Frontiers in Artificial Intelligence* 4: 648583. doi 10.3389/frai.201.648583.

Zhu, J. & Jürgens, D. 2021. The structure of online social networks modulates the rate of lexical change. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 2201–2218, Online. Association for Computational Linguistics.

Studying digital communication of multilingual communities: how to strive towards sustainability in CMC studies?

Martti Mäkinen
Hanken School of Economics
PO Box 479, 00101 Helsinki, Finland
martti.makinen@hanken.fi

Abstract

This paper discusses the sustainability of linguistic research of CMC. The paper argues for contextualisation of the linguistic phenomena studied and the role of informant and researcher identities in any investigation, motivating this proposition through the concept of language ecology. The arguments are exemplified through the author's research project on Finnish Swedish speaking young adults' instant messaging discussions on WhatsApp.

Keywords: language ecology, multilingualism, sustainability, polylingualism, languaging, code-switching, Finnish-Swedish

1. Introduction

Sustainability is often defined through the dimensions of environmental, social, and economic sustainability, offering means to meet the present needs of humanity without jeopardizing the ability of future generations to meet their needs (Brundtland, 1987). If the dimensions of sustainability are depicted as a Venn diagram (see Figure 1), sustainability "happens" in the area where the three dimensions meet. Of late, the dimension of *culture* has been often added as the fourth dimension of sustainability, and as any of the other dimensions, it cuts across the three others described above.

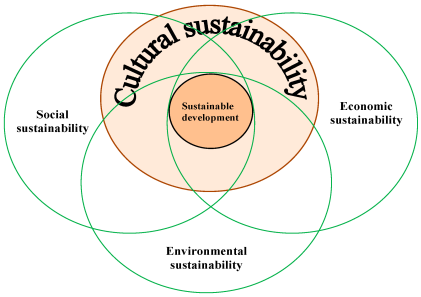


Figure 1: The four dimensions of sustainability (Source: Pop et al., 2019).

Sustainability in and of language studies has been a popular concept in recent years, and it is only natural that it is extended to the study of CMC as well. In the current paper, we exemplify the concept of sustainability in the study of languages through our work on a CMC corpus in preparation, *Multilingual / Multimodal WhatsApp Discussions Hanken* (henceforth *MMWAH*). The langugage data collected for the corpus is from Swedish speakers living in Finland, and as the other official language of the country, Finnish, happens to be the majority language, the speakers Swedish (or Finnish Swedish) are also speakers of a minority language.

Sustainability here will be investigated through the concept of *language ecology* (cf. Haugen, 2001),[1] which can roughly be defined as the interaction between a language and its context. In multilingual communication, the context of a language consists of other languages, and ignoring that environment means ignoring such elements in the language that come about through interaction with other languages (cf. Bastardas-Boada, 2014). In MMWAH data, the study of English is the study of the language embedded in other languages, in this case (often) in Swedish and Finnish. As the informants of MMWAH are mainly multilingual, part of the environment with which their mother tongue interacts is formed by one or more languages spoken by the same speakers. For multilingual language users and communities, sustainability bears a particular relevance as the relationship between the languages in use often define one or more of them as minority languages.

2. About the project

Project *Multilingual Instant Messaging: Focus on WhatsApp in Finnish-Swedish Digital Communication* is a three-year project funded by the Swedish Cultural Foundation in Finland begun in August 2022. The main aim of the project is to document, chart, and investigate the multilingual and polysemiotic practices in CMC of Swedish speakers in Finland.

WhatsApp has been chosen as the instant messaging platform of interest, due to its wide use in Finland. According to AudienceProject (2020) and DNA (2022), WA is the most used IM platform in Finland, with 87 % of the population having some experience in its use, with 62 % of the population between 16–74 years of age using the app daily. WhatsApp is geared towards the ease of textual communication, and that feature combined with the overall popularity decided the choice over other platforms that may be equally or more popular among young adults, e.g. Discord, Snapchat, etc. As WhatsApp

[1] Language ecology in association with CMC has been studied before, see e.g. Coats (2017), but the consideration of sustainability with CMC and language ecology is a rarer case in earlier literature.

is an unmoderated platform, it provides an environment through which spontaneous and authentic language use can be observed. Therefore we also think that WhatsApp data will allow us to observe linguistic innovations as they happen (Mäkinen, 2023).

The project collects WhatsApp data from young adult speakers of Finnish Swedish (between 18–30 years), and the linguistic circumstances in Finland (in addition to schools teaching compulsory courses in both official languages) have made bilingualism a practical choice for many. Adding to this the ubiquitousness of English in media, education, and working life, many Swedish speakers in Finland are, *de facto*, trilingual. Therefore, the data collected for MMWAH attest to the use of not only Swedish, but both domestic languages (Fi. and Sw.) and English.

3. Sustainability in language studies

Relevant questions to be presented to MMWAH corpus are, how do the three expected languages co-exist in CMC in Finland, in what manner is the dynamic interplay between the languages manifested, and how do user groups build identities through the available linguistic and other semiotic repertoires (cf. Peterson et al., 2022). These questions all bear relevance to the dimensions of sustainability, and not least to the social and cultural dimensions. In this section, the different factions and actors related to the study of CMC will be discussed.

3.1. Language users of today and tomorrow

Speakers of Finnish Swedish are known to mix languages in their social media posts, as in examples (1) and (2):

(1) *Joo let's tehään niin!!*
Yes let's $\text{do}_{\text{passive}}$ that
'Yes, let's do that!'

Example (1) uses elements from Swedish, English and Finnish. "Joo" is (possibly) a Finnish rendering of Swedish "Jå(å)" (meaning "yes"), and "tehään niin" is dialectal Finnish for "let's do it", making the English "let's" somewhat redundant.

(2) *Jag tror att vi kan ba fa dit för det*
I think that we can just go there as it
e int anymore nån wanhojentanssi tyyppei
is not anymore any prom guys
'I think we can just go there as there will be no-one (looking for a) prom (dress) anymore.'

In (2), we see again a mix of three languages, the English "anymore" filling in for the Sw. word "mera" in the (dialectal) Swedish part of the sentence, the last two words of the sentence being in dialectal Finnish.

The examples above can be said to attest to *language mixing*, and maybe also *code-switching* (Franceschini, 1998, p. 51). Nevertheless, according to some definitions (Meakins, 2018), the former term would mean creating a new code while there exists a (named) language code mutually intelligible to all discussion participants; hence, the mixed code would only serve the creation of group identity, without further communicative purposes. This seems unlikely: even phatic expressions are communication. The latter term, code-switching, would require at least two distinct named languages, of which one functions as the matrix languages in which switches in codes are embedded. Even this does not seem to describe the situation satisfactorily.

The examples above may also be cases of *plurilingualism* (García and Wei, 2015; García and Otheguy, 2020), or even *polylanguaging* (Jørgensen, 2008; Jørgensen and Møller, 2014, p. 69). Here, the chat participants appear to be involved in the creation and use of one consistent code, drawing on their full linguistic repertoire without concern for for any social or political constraints in their use (Otheguy et al., 2015). The linguistic resources consist of elements from three different languages (that outsiders can name) and several dialectal variants thereof.

From the point-of-view of an individual language user, it may be significant that the kind of usage exemplified in (1) and (2) is recorded: we are looking at unique and idiosyncractic use of language, and there is value in recording it and preserving it. Our interest in this variant of language mixing vindicates the original language users. This idea of value in everyday, non-standard language use complies with the social, environmental, and cultural dimensions of sustainability. Research in unmoderated social media platforms records a "voice" that would go unnoticed in linguistic research unless this particular effort is made. As such platforms very likely are seats of linguistic innovations, it would be a great pity not to observe them while we can.

3.2. Contemporary and future researchers

For the examples (1) and (2), *languaging* is a neutral enough term for analysing the chat situation, as it does not bind researchers' hands to a certain framework before decisions on the kind of multilingual practices have been made. For a researcher, a point of sustainability is not be trapped by terminology and related theoretical framework before the data has informed them in the choice of both.

The above requires accepting the dual nature of languages: We can identify a language by name, (a named language), which is a 'monolithic' entity, has a standard that is taught and maintained, and can often be associated with a state and / or a nation. At the same time we need to accept the idea of language as a process:

> There is no dispute about the fact that languages are socially constructed with porous boundaries, but languages are also experientially and socially real for students, teachers, policymakers, curriculum designers, politicians, and most researchers [. . .] there is no contradiction between treating languages as both processes and, at the same time, as concrete entities. Cummins, 2021

Investigators must also question their own identity and mandate as researchers. In the MMWAH project, the current author is a Finnish-speaking scholar of English with twice the age of the target group, hence there is no mandate to study the Finnish Swedish of young adults, not professionally, nor socially. The bottom line is that data donors must be able to trust the research project with their data, and that trust is created through respect towards the informants and the data, relevance of the study, reciprocity ("What's in it for the data donors?"), and responsibility, i.e. taking care of donor identities and the data (Kosner, 2023).

Let us still consider briefly the needs of contemporary and future researchers. GDPR protects the privacy of participants in research so that the collected data originating with natural persons will not allow direct or indirect identification of the persons in question (European Parliament, 2016). Compliance with GDPR requires researchers to take such measures that identification of research informants is impossible under any circumstances. To guarantee this, every HEI have devised data management plan templates that inform researchers in this. However, DMPs can also be restrictive, and templates that propose deletion of data after a certain number of years are real, even if they are based on a clear misinterpretation of the law. The idea of deleting data after the completion of a project may arise from GDPR articles 5(1)(c) (data storage limitation) and 5(1)(e) (data minimisation); however, these articles limit the storage and collection of *personal* data, hence anonymised data can be stored indefinitely.[2] All in all, deleting data would be waste of resources and effort, and it would deny future researchers an interesting tool for other, synchronic and diachronic research questions.

Finally, an often discussed notion is complying with the FAIR principles (findability, accessibility, interoperability, and reusability) (Frey et al., 2019), which are in accordance with at least economic and social, and possibly also environmental sustainability.

3.3. Policy makers, language ecology

In examples (1) and (2), we may be far away from *parallellingualism* (Hultgren, 2014, pp. 68–69; Harder, 2008); nevertheless, research on social media language use informs our policy makers. Current discussions on the influence of English on domestic languages seem to leave no-one cold. In Finland, in November 2023, the results of Finnish government's survey on the role and position of English in the Finnish society were published (Laitinen et al., 2023), and they again fanned the public debate about the alleged threat English poses on the domestic languages. Investigations into CMC will provide useful and up-to-date information on concurrent language use, and that provides a reliable basis for any decisions on language policies and legislation. Such relevance of research is also sought after by local funding bodies: bringing research closer to home is an argument that sides with ideas about economic and environmental sustainability.

The coexistence of several languages in MMWAH data underlines the organic nature of communication: it is always purposeful and practical, and rather than seeking to flout standards it seeks to be effective, understandable, and simultaneously individual and communal. Studies of languages in CMC in Finland will make apparent how domestic languages, English, and other languages cohabit in the minds of language users in Finland, and therefore these languages should not be ignored when decisions about policies encoding the statuses of languages are made. Research in CMC is, thus, a valuable tool in the societal discussion on the ecology of languages in which we communicate, and it will have repercussions on current and future communicative cultures and on social and cultural sustainability in general.

4. Conclusion

In the light of this paper on sustainability and the study of digital communication in multilingual communities, we arrive at the following list of (more or less generalisable) key take-aways:

- Serve the academic community and the society.
- Know yourself, the informants, potential audiences, the data, the societal powers at play, and the needs of your colleagues.
- Waste not, want not (that is, data).
  - Contextualize everything.
  - Let data guide your choice of terminology.
  - Make data accessible and reusable.
- Accept the dual nature of languages.
- Be mindful of people, data, and results.

5. Acknowledgements

The project Multilingual Instant Messaging: Focus on WhatsApp in Finnish Swedish Digital Communication wishes to acknowledge the support of Swedish Cultural Fund in Finland, 2022-2025.

6. Bibliographical References

AudienceProject (2020). *Insights 2020. App & Social Media Usage*. Ed. by Rune Werliin. URL: https://audienceproject.com/wp-content/uploads/AudienceProject_Study_App_and_Social_Media_Usage_2020_pdf.pdf?x17261.

Bastardas-Boada, Albert (2014). "Linguistic Sustainability for a Multilingual Humanity". In: *Sustainable Multilingualism / Darnioji Daugiakalbystė* 5, pp. 134–163.

[2] In MMWAH project, anonymisation is taken more than seriously: the target group of potential data donors is only 40,000 strong, therefore it is not enough to anonymise names and place names, but one needs to edit also circumstantial evidence in the data in order to prevent unintended identification due to the small, tight-knit language user community.

Brundtland, Gro (1987). *Our common future: Report of the World Commission on Environment and Development*. Oxford University Press.

Coats, Steven (2017). "European Language Ecology and Bilingualism with English on Twitter". In: *Proceedings of the 5th Conference on CMC and Social Media Corpora for the Humanities (cmccorpora17)*. Ed. by Ciara R. Wigham Egon W. Stemle. Bolzano: Eurac Research, pp. 35–38.

Cummins, Jim (2021). "2. Translanguaging: A Critical Analysis of Theoretical Claims". In: *Pedagogical Translanguaging: Theoretical, Methodological and Empirical Perspectives*. Ed. by Päivi Juvonen and Marie Källkvist. Bristol: Multilingual Matters, pp. 7–36. DOI: 10.21832/9781788927383-004.

DNA (2022). *DNA Digitaaliset elämäntavat – tutkimus*. URL: https://www.dna.fi/documents/753910/11433306/Digitaaliset_elamantavat_tutkimusraportti_2022.pdf/.

European Parliament (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. https://gdprinfo.eu/eu/.

Franceschini, Rita (1998). "Code-Switching and the Notion of Code in Linguistics". In: *Code-Switching in Conversation: Language, Interaction and Identity*. Routledge. Chap. 4, pp. 223–240. DOI: https://doi.org/10.4324/9780203017883-4.

Frey, Jennifer-Carmen, Alexander König, and Egon W. Stemle (2019). "How FAIR are CMC corpora?" In: *Proceedings of the 7th Conference on CMC and Social Media Corpora for the Humanities (cmccorpora19)*, pp. 25–30.

García, Ofelia and Ricardo Otheguy (2020). "Plurilingualism and translanguaging: commonalities and divergences". In: *International Journal of Bilingual Education and Bilingualism* 23.1, pp. 17–35. DOI: 10.1080/13670050.2019.1598932.

García, Ofelia and Li Wei (2015). "Translanguaging, Bilingualism, and Bilingual Education". In: *The Handbook of Bilingual and Multilingual Education*. Ed. by O. García W. E. Wright S. Boun. John Wiley & Sons Ltd, pp. 223–240. DOI: 10.1002/9781118533406.ch13.

Harder, Peter (2008). *Hvad er parallelsproglighed?* URL: http://cip.ku.dk/om_parallelsproglighed/oversigtsartikler_om_parallelsproglighed/hvad_er_parallelsproglighed/.

Haugen, Einar (2001). "The Ecology of Language". In: *The ecolinguistics reader: language, ecology and environment*. Ed. by Alwin Fill and Peter Mühlhäusler. London: Continuum, pp. 57–66.

Hultgren, Anna Kristina (2014). "Whose parallellingualism? Overt and covert ideologies in Danish university language policies". In: *Multilingua* 33.1-2, pp. 61–87. DOI: 10.1515/multi-2014-0004. URL: https://doi.org/10.1515/multi-2014-0004.

Jørgensen, Normann (2008). "Polylingual Languaging Around and Among Children and Adolescents". In: *International Journal of Multilingualism* 5.3, pp. 161–176. DOI: 10.1080/14790710802387562.

Jørgensen, Normann and J. Møller (2014). "Polylingualism and Languaging". In: ed. by C. Leung & B. V. Street. The Routledge Companion to English Studies. Routledge, pp. 67–83. DOI: 10.4324/9781315852515.

Kosner, Lukas (2023). *Sustainability in linguistics: a mirror image?* URL: https://blogs.helsinki.fi/linguisticsandsustainability/2023/10/03/sustainability-in-linguistics-a-mirror-image/.

Laitinen, Mikko, Sirpa Leppänen, Paula Rautionaho, and Sara Backman (2023). *Kohti joustavaa monikielisyyttä*. Helsinki: Valto.

Mäkinen, Martti (2023). "MMWAH! Compiling a Corpus of Multilingual / Multimodal WhatsApp Discussions by Swedish-speaking Young Adults in Finland". In: *Proceedings of the 10th International Conference on CMC and Social Media Corpora for the Humanities*. Ed. by Louis Cotgrove, Laura Herzberg, Harald Lüngen, and Ines Pisetta. Leibniz-Institut für Deutsche Sprache, pp. 136–139. DOI: 10.14618/1z5k-pb25.

Meakins, Felicity (July 2018). "Mixed languages". In: DOI: 10.1093/acrefore/9780199384655.013.151.

Otheguy, Ricardo, Ofelia García, and Wallis Reid (2015). "Clarifying translanguaging and deconstructing named languages: A perspective from linguistics". In: *Applied Linguistics Review* 6.3, pp. 281–307. DOI: 10.1515/applirev-2015-0014.

Peterson, Elizabeth, Turo Hiltunen, and Johanna Vaattovaara (2022). "A Place for pliis in Finnish: A Discourse-Pragmatic Variation Account of Position". In: *Discourse-Pragmatic Variation and Change: Theory, Innovations, Contact*. Cambridge: Cambridge University Press, pp. 272–290.

Pop, Izabela Luiza, Anca Borza, Anuța Buiga, Diana Ighian, and Rita Toader (2019). "Achieving Cultural Sustainability in Museums: A Step Toward Sustainable Development". In: *Sustainability* 11.4. DOI: 10.3390/su11040970.

Spatial and Temporal Deixis in digital asynchronous discussions: where is “here”, when is “now”?

Michel Marcoccia
Université de technologie de Troyes, LIST3N (Laboratoire Informatique & Société Numérique)
E-mail: michel.marcoccia@utt.fr

Abstract

This article analyzes the way in which spatial and temporal deixis is constructed in a corpus of 400 messages extracted from a French-speaking discussion forum. The analysis of the occurrences of "here", "now", "tomorrow", and "yesterday" allows us to study the way the discussants construct this deixis in a communication situation marked by spatial and temporal disjunction.

Keywords: spatial, temporal, deixis, discussion forum

1. Introduction

Deixis concerns the use of linguistic items which rely on the context for their interpretation. Spatial deixis involves the specification of locations relative to points of reference in the speech event. Temporal deixis refers to the time in which an event takes place. It involves the use of temporal expressions such as "*now*", "*then*", "*yesterday*", etc.

Digital technologies increase our communication capacity by providing new communication channels, but at the same time they create interaction contexts far removed from the model of face-to-face interaction. Digital media separate the place and the time of message production from that of its reception.

Regarding the spatial framework, a consequence of this disjunction is to lead users to simultaneously experience two distinct environments: the physical environment in which the user is actually present and the environment constituted by the digital discussion platform. This experience of a digital media space corresponds to what is called “*telepresence*”, that is to say the perception of an space for interaction space, materialized by a technical platform and constituted by the digital conversations that take place there.

Regarding the temporal framework, asynchronous digital communication is a sort of paradox: the temporal framework is vague to the extent that the "*now*" of the author of a message does not correspond to the "*now*" of the reader but it is at the same time very clear because the digital platform allowing online discussions generally date the messages, so that each message corresponds in principle to a very precise temporal referent.

Thus, participation in an online discussion, in a discussion forum or a site for commenting on newspapers articles, implies that the participants experience two spatial environments at the same time: the spatial environment in which they are physically present (at home, in their office...) and the digital environment built by the platform used for the online discussion. They also experience a temporal framework based on the disjunction between the time of production of a message (precisely indicated by the system) and the time of its reception.

From a theoretical point of view, this work amounts to questioning the notion of spatial and temporal deixis in a context of digital communication. The question is: from what reference will spatial and temporal deixis be constructed in digital discussions?

This article takes up and extends more general work on the analysis of written digital communication and on the problem of deixis in digital discourse (Marcoccia, 2016; 2023).

2. Corpus and methodology

The present study consists of analyzing 400 messages, with 100 occurrences of the terms “*ici*” (“*here*”), “*maintenant*” (“*now*”), “*hier*” (“*yesterday*”) et “*demain*” (“*tomorrow*”), taken from messages posted in several sub-sections of the discussion forums on the *Doctissimo* site.

Doctissimo is a French-speaking website dedicated to health and well-being. It is one of the sites with the largest audience in these areas. This site offers access to various discussion forums.

The representativeness of the phenomena observed is not based on statistics but on the knowledge of the discussion forums by the researcher, who combines the methods of persistent observation (Herring, 2004) and “small corpus” analysis (Danino, 2018).

The research method unfolds in four phases, progressing from persistent observation to a final phase involving pragmatic-semantic analysis of the referents of the deictic term “ici” (“here”). In each phase, we focused on: (1) Selecting sub-forums: “Animals / Veterinary Medicine” and “Coronavirus / Vaccines.” These forums were chosen for their diversity and their representativeness of typical forum usage on Doctissimo, including mutual social support, self-disclosure, and narratives (Gauducheau & Marcoccia, 2011). (2) Utilizing the keyword search tool on the Doctissimo forum platform to extract messages containing the term “ici” (“here”) for spatial deixis, as it is arguably the most direct and universal example of spatial deixis (Diessel 1999: 38), as well as “hier / demain / maintenant” (“yesterday / tomorrow / now”) for temporal deixis. (3) Compiling a corpus of 100 occurences. (4) Identifying the various referents of these deictics (“here”)

within the messages to perform a comparative analysis of the different ways space and time are discursively constructed in these discussions. The interpretation of deictics in the messages is based on the researcher's understanding and how response messages reveal users' interpretations of these deictics, adhering to the principle of dialogic interpretation.

3. Spatial deixis : “ici” (here)

Spatial deixis pertains to a principle of location crucial in discourse exchanges, as it anchors participants to a specific context that imparts meaning to their interactions (Cairns 1991: 19). Traditionally, spatial deixis is viewed as more “basic” than temporal deixis (Lyons, 1977: 669). In the realm of spatial deictics, “ici” (“here”) is typically seen as a “transparent” deictic since its referent is clear and inherently tied to the speaker's location during the act of communication (Kleiber, 1986: 5). However, digital communication complicates the reference of this deictic. What serves as the spatial reference point in such contexts? Linguists agree that a deictic expression is interpreted based on the extralinguistic context of the utterance. Yet, this definition becomes problematic in online communication. In online discussions, what constitutes the extralinguistic context? Is it the physical locations of the speakers or the online platform facilitating their interaction?

From a pragmatic standpoint, Yule (1996: 9) states that the primary distinction in deixis is “near the speaker” versus “far from the speaker.” Thus, there are proximal terms (this, here, now) and distal terms (that, there, then), differentiating what is close to or far from the speaker. Proximal terms generally refer to the speaker’s location, the deictic center, hence “here” and “now” typically denote the speaker’s place and time. This proximal/distal distinction does not necessarily apply to online discussions. Participants in a discussion forum are physically distant yet close in the sense that they share the same platform for interaction.

In essence, digital communication challenges the traditional definitions of spatial deixis. This situation diverges significantly from the canonical context in which deictics are typically interpreted. Lyons (1977: 367) describes the canonical context as involving a shared physical and temporal setting (“here” is also “now”), based on the *origo*—the deictic center defined by “I,” “here,” and “now.” In this model, the speaker at the moment of utterance serves as the referent and anchor for personal, spatial, and temporal orientation, organized in an “egotic” manner (Levinson, 1983: 63). However, online asynchronous discussions, like those in forums, deviate from this canonical context: they lack a stable “I,” “here,” and “now.”

In fact, digital communication is not the sole context where spatial deixis deviates from the canonical model (Bazzanella, 2019: 7). Recent studies reveal that “here” functions more complexly than classical analyses suggest, encompassing spatial, non-spatial, temporal, and textual uses (where “here” refers to a location within a text). The uses of “ici” (“here”) are far more varied than traditionally recognized (Kleiber, 2008, 2018 ; Le Draoulec & Borillo,). Moreover, many studies discuss “deictic projection,” where the spatial origin extends to peripheral uses of spatial deictics, such as a place, city, nation, imaginary place, speaker’s body, visual path, or another person’s location as a reference point. In these instances, context is crucial for understanding spatial deictics beyond the canonical situation (Bazzanella, 2019: 7).

The issue of space in digital communication has rarely been explored linguistically, particularly concerning deixis. Two noteworthy works are Holmes’ 1995 paper and Dostalek’s 2020 Master’s thesis. Holmes’ paper examines how participants in computer-mediated communication (CMC) adapt their descriptions of location and spatial relationships to online conversation constraints, emphasizing language’s role in telepresence. The study analyzes deixis in synchronous digital discussions and shows that participants use both physical and network locations as reference points, with digital space being the primary frame of reference. This work highlights the adaptability of deictic language and spatial reference in action. Our chapter aligns with Holmes' perspective (1995). Dostalek’s 2020 thesis highlights the limited use of spatial deixis in asynchronous forums due to the lack of a shared physical context. The findings of this chapter do not align with Dostalek’s observations. Indeed, “here” is present in many messages.

It is interesting to study the way in which participants in online discussions refer to their spatial locations and, in particular, to analyze the contrast between those who refer to their physical spatial locations and those who highlight their “virtual” presence in the shared digital discussion space. Here, this phenomenon is studied in a restrictive way, in analyzing one of its most visible manifestations: the use of the French spatial deictic expression “*ici*” (“*here*”) in messages posted in various Doctissimo sub-forums.

Through identifying occurences of the deictic "*ici*" ("*here*"), a comparative analysis was conducted to discern various modes of spatial discourse construction within these digital discussions.

Our findings show a stark contrast between three cases:

1. "*ici*" ("*here*") functions as a “techno-word” (Paveau, 2015), a term associated with a navigation function in the hypertext that constitutes the Internet network (14%) : *Le mal continue ? Si non tu peux cliquer ici pour voir des solutions (The pain continues? If not you can click here to see solutions)*.
2. “*ici*” denotes the sender's physical context (35%) : *Ben il semblerait que peu d’écoles le respectent, en tout cas par ici* (*Well it seems that few schools respect him, at least around here)..*
3. “*ici*” signifies the digital discussion space (51%) : *N’ayant pas l’habitude de ce type de forum, je suis venu ici, car je suis un peu désemparé et j’ai l’impression de plus avoir ma vie d’avant depuis désormais 4 mois.. (Not used to this type of forum, I came here because I’m a little confused and I feel like I no longer have my life before for 4 months now...)*.

This dichotomy between the cases (2) and (3) can shed light on the presence or absence of a shared spatial consciousness or community affiliation.
Furthermore, this study delineates the diverse boundaries of the digital space marked by "*ici*" ("*here*"), ranging from the broad expanse of "cyberspace" to specific discussion threads. Similarly, employing "*ici*" ("*here*") to denote physical space can contribute to a spectrum of identity construction strategies, spanning from national affiliations ("ici = in France") to individualistic expressions (when "ici" represents "me").
It is interesting to note that certain uses of "here" described in the literature are absent in this corpus, such as abstract or notional uses.
This study shows how individuals in online written discussions skillfully adjust their spatial references to fit the specific dynamics of digital communication.

4. Temporal deixis : “hier” (yesterday), “maintenant” (now), “demain” (tomorrow)

On the linguistic level, temporal deixis is constructed using verb tenses and a series of markers (adverbs: “*yesterday*”, “*tomorrow*”, noun phrases: “*last week*”, prepositional phrases: “*in two months*”). On the semantic level, temporal deictics are classified in relation to the moment of utterance, in terms of anteriority (“*yesterday*”), simultaneity (“*now*”), posteriority (“*tomorrow*”), neutrality (“*today*”) or indifference (“*just now*”) (Ticca, Traverso & Ursi, 2017).
Different temporal expressions can be distinguished (Moeschler, 1991):
- referentially autonomous temporal expressions, defined (which indicate a precise temporal reference point like “*September 7, 1966*”) or indefinite (like “*one day*”), which allow, in a non-indexical way, to fix a point of temporal reference in relation to which the other temporal marks (verbal tenses in particular) will set their temporal reference.
- Non-referentially autonomous temporal expressions, which constitute all anaphoric and deictic temporal expressions. Anaphorics need to be in a coreference relationship (partial or total) with a referentially autonomous temporal expression. Deictics (“*now*”, “*today*”, “*yesterday*”, “*tomorrow*”, “*next week*”, etc.) are characterized in that their reference is variable depending on the moment of utterance.
Discussion forum is an obvious example of asynchronous written communication, which implies that writers and readers do not share the same time frame. In theory, if we consider that participants in the discussion try to make their messages as intelligible as possible, purely deictic temporal expressions should be used less than other types of temporal expression. From this point of view, we could expect that written digital communication would show the typical type of use of deicitics as written communication (with less deictics than oral, see Biber et al. 1999, Skogs 2014). In fact, as Skogs (2014) shows, we can observe numerous deictics in texts produced in digital written communication situations. It is interesting to see how this situation is handled by participants in an online discussion, in particular, if they use "*hier / yesterday*", "*demain / tomorrow*", or "*maintenant / now*" whereas the situation is supposed to make these deictics difficult to interpret.
The analysis of a corpus of 100 occurrences of "*hier / yesterday*", "*demain / tomorrow*" and "*maintenant / now*" in messages sent to the Doctisissmo forum shows that the use of these deictics seems rather common and that they are found in messages in which no complementary temporal reference is provided to allow the interpretation of the temporal reference carried by the deicitic.
In other words, using “*hier / yesterday*”, “*maintenant / now*” or “*demain / tomorrow*” in a message posted in this forum is almost never a problem. In only one case, a user highlights the lack of temporal synchronization between messages by ironically pointing out to another user that he is responding to a message posted several years previously.
Two hypotheses can be put forward to explain this use, which is nevertheless risky from the point of view of message intelligibility.
- The system automatically dates messages and this date appears at the top of these messages. This temporal reference allows the deictic temporal expressions to be understandable because they are co-referential.
- The participants refer only to the temporal framework in which they are located and act as if the readers of their messages shared this temporal framework.
In fact, our hypothesis is that the digital discussion creates a sort of illusion of shared temporality which leads certain users to act as if the exchanges were synchronous or quasi-synchronous, for example by responding in February 2024 to a message posted in 2018.

5. Conclusion

Various lessons can be drawn from this study.
Talking about reference to the context to specify the meaning of the spatial and temporal deictics used in discourse and interaction is not necessarily relevant in compter-mediated communication, since the context is not given a priori and it does not become a reality shared by the participants only through a process of interactional construction.
In our study, we observe that spatial deixis can refer to two very different spatial frameworks: the disjointed physical spaces of the participants in the discussion or the shared discussion space which can serve as a frame of reference from which the spatial deictics are produced and interpreted.
In the same way, the temporal deixis in the digital discussions of our corpus seems to be constructed on the basis of a constructed synchrony. The asynchronous nature of the discussion is neglected and the participants choose to use deictic temporal expressions that are not referentially autonomous even though they do not share the same temporal frame of reference. In a way, we can consider that, in a discussion forum, temporal deixis is built on the idea that time is suspended.
Thus, we can analyze these results as testifying to the fact that the participants feel engaged in an interaction situation

in which they share the same spatio-temporal framework. This illusion seems to us to be quite constitutive of the specificities of online discussions which always seem to have the face-to-face conversation as a frame of reference. To extend this work, various questions can be developed. First of all, it would be interesting to check if there is a relationship between the different uses of deicitics and the illocutionary or argumentative value of the messages in which they appear. Thus, "here" can help strengthen an argument by example. Furthermore, the construction of a shared space and temporality is incontestably linked to the community character of the forum studied in this work. An analysis comparing these results to other forums would make it possible to move forward on this question.

6. References

Bazzanella, C. (2019). The Complex Process of Mis/Understanding Spatial Deixis in Face-to-Face Interaction., *Pragmática Sociocultural / Sociocultural Pragmatic* 7: pp.1-18.

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

Cairns, B. (1991). Spatial Deixis. The Use of Spatial Co-ordinates in Spoken Language. *Lund University, Department of Linguistics, Working Papers* 38, pp.19-28.

Danino, C. (2018). Les petits corpus. Introduction., *Corpus* 18. http://journals.openedition.org/corpus/3099

Diessel, H. (1999). *Demonstratives: Form, function and grammaticalization*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Dostalek, T. (2020). *Reference and deixis in internet forums*. Thesis for Master's degree, Univerzita Pardubice, Czechia.

Gauducheau, N., Marcoccia, M. (2011). Le soutien social dans les forums de discussion Internet : réalisations interactionnelles et contrat de communication. In P. Castel, E. Salès-Wuillemin & M.-F. Lacassagne (eds.), *Psychologie sociale, communication et langage. De la conception aux applications,* Bruxelles: De Boeck Editions, pp.349-368.

Herring, S. C. (2004). Computer-Mediated Discourse Analysis: An Approach to Researching Online Communities. In S.A. Barab, R. Kling, and J. H. Gray (eds.), *Designing for Virtual Communities in the Service of Learning*, Cambridge: Cambridge University Press, pp. 338-376.

Holmes, M.E. (1995). Naming virtual space in computer-mediated conversation, *ETC: A Review of General Semantics* 52 (2): pp.212-221.

Kleiber, G. (1986). Déictiques, embrayeurs, "token-réflexives", symboles indexicaux, etc. : comment les définir ?, *L'Information Grammaticale* 30: pp.3-22.

Kleiber, G. (2008). Comment fonctionne *ICI, Cahiers Chronos* 20: pp.113-145.

Kleiber, G. (2018). *Ici* en glanures, *Langue Française* 197: pp.35-49.

Le Draoulec, A, Borillo, A. (2013). Quand *ici*, c'est *maintenant*, Langue française 179: pp.69-87.

Levinson, S.C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Lyons, J. (1977). *Semantics 2*. Cambridge: Cambridge University Press.

Marcoccia, M. (2016). *Analyser la communication numérique écrite*, Malakoff: Armand Colin.

Marcoccia, M. (2023). The physical-digital interface. What does "ici" ("here") mean in a written online discussion?, In A. Parini & F. Yus (eds.), *The Discursive Construction of Place in the Digital Age*, New York: Routledge, pp.152-168.

Moeschler, J. (1991). Référence temporelle et déixis: vers une approche milnérienne. *Travaux neuchâtelois de Linguistique*, (17), pp.97–122.

Paveau, M.-A. (2015). Ce qui s'écrit dans les univers numériques. Matières technolangagières et formes technodiscursives, *Itinéraires LTC* 2014-1. http://journals.openedition.org/itineraires/2313

Skogs, J. (2014). Features of Orality, Academic Writing and Interaction in Asynchronic Student Discussion Forums. *Nordic Journal of English Studies* Vol. 23(3): pp. 54-82.

Ticca, A., Traverso, V. & Ursi, B. (2017). Multidimensionnalité, indexicalité et temporalité(s) : le cas de *tout à l'heure*, *Langue française* 193, pp.57-76

Yule, G. (1996). *Pragmatics*. Oxford: Oxford University Press.

Gay slang on Facebook: Subversion or stigmatization?

Yonatan Marik & Hadar Netz
Program for Multilingual Education, School of Education, Tel Aviv University
E-mail: yonatan.marik@gmail.com, hadar.netz@gmail.com

Abstract

Social Networking Sites (SNSs) offer users increased visibility and connections, serving as supportive avenues for self-expression and social bonding. Despite these benefits, SNSs also pose risks, particularly for marginalized communities. This study addresses the gap in research on LGBTQ+ individuals' discursive practices on these platforms by examining the use of *oxtšit*, a gay Hebrew slang, within a gay, Hebrew-speaking community on Facebook. Investigation focuses on the functions of grammatical gender alternations between masculine and feminine forms. Findings suggest that while speakers no longer distance themselves from the *oxtša* persona, they still do not fully embrace it. Specifically, while freely using the *oxtša* persona to describe stereotypically feminine emotions like love and romance, the persona remains excluded from descriptions of erotic, sexual desires and acts. Group members thus render the *oxtša* persona sexually unattractive and undesirable, thereby reinforcing heteronormative masculinity at least to some extent.

Keywords: social networking sites, Facebook, LGBTQ+, Camp, slang, grammatical gender alternations

1. Introduction

From its inception, the Internet has been seen as a forum promoting democratic engagement, with the potential to break down social barriers and amplify marginalized voices (Dori-Hacohen & Shavit, 2013). Social Networking Sites (SNSs), such as Facebook, Instagram, and others, indeed provide users with various affordances, including increased visibility and connections that surpass social hierarchies and geographical barriers (Hiebert & Kortes-Miller, 2023; Fox & Ralston, 2016). These platforms offer supportive avenues for self-expression, fostering social bonds, and facilitating learning (Fox & Ralston, 2016), potentially contributing to one's resilience and well-being (Hiebert & Kortes-Miller, 2023). Such affordances are particularly significant for marginalized communities, such as the LGBTQ+ community, which often faces ongoing stigma, discrimination, and bullying (Hiebert & Kortes-Miller, 2023; Fox & Ralston, 2016; Marciano, 2010).

However, SNSs also come with limitations and potential risks for their users, especially those from marginalized communities. For instance, limited visibility control and *context collapse* (boyd, 2011) increase the risk of unintended exposure (boyd, 2011; Duguay, 2016). Furthermore, certain SNSs, such as Facebook, enforce a real-name policy (Fox & Ralston, 2016), eliminating the option of anonymity available on other platforms. Context collapse and the *nonymous* nature of platforms like Facebook (Fox & Ralston, 2016) can pose significant risks, particularly for LGBTQ+ individuals, who may not be equally or entirely open about their identity with all their family and friends (DeVito et al., 2018). Lastly, while participating in online communities presents opportunities to connect with others who share similar experiences, it also carries the heightened risk of individuals experiencing online discrimination and bullying firsthand, as well as the risk of being exposed to challenging experiences of others within these communities. Research indicates that these risks may lead to increased anxiety and reluctance to further engage (Fox & Ralston, 2016).

This underscores the significance of researching online communities, especially those belonging to marginalized groups such as the LGBTQ+ community. While some studies on online LGBTQ+ communities exist, there has been limited focus on the discursive practices of LGBTQ+ individuals on online platforms, despite the central role of self-expression. The present study addresses this gap by examining a specific case: the use of *oxtšit*, a gay Hebrew slang (Brom, 2022; Levon, 2012), akin to *verbal Camp* (Harvey, 1998; Vider, 2013), within a gay, Hebrew-speaking community on Facebook.

2. *OXTŠIT* in social media times

Oxtšit is a Hebrew slang variety associated with the gay Hebrew-speaking community in Israel, especially with feminine gay men and transexual women (Brom, 2022). The term *oxtšit* comes from the word *oxtša*, derived from "my sister" in Arabic, and is used within this community to describe gay Middle Eastern men who embrace feminine characteristics, such as wearing makeup and designer clothes, and prefer to take the passive role during sexual intercourse (Levon, 2012). This slang variety has its own lexicon (Levon, 2012) as well as its own phonology, morphology, and syntax (Brom, 2022), thus aligning with the concept of a speech community by virtue of its close association with the speakers' identities (Hall & Nilep, 2015).

Levon (2012) originally described *oxtšit* as a "secret lexicon," consisting of only 28 entries primarily used within institutionalized gay and lesbian social forums (p. 189). However, a decade later, Brom (2022) argued that *oxtšit* is no longer confined to secrecy, having entered mainstream culture, likely through mass media. Additionally, Brom (2022) posits that *oxtšit* encompasses various linguistic aspects beyond lexical features, including phonology, morphology, and syntax. For instance,

phonologically, *oxtšit* adopts the non-voiced fricative guttural [ħ] for the phoneme /x/, influenced by Arabic and *Mizrahi* Hebrew phonology (ibid.). Morphologically, it employs suffixes like [-aʒ], often replacing the standard Hebrew masculine plural suffix [-im] (ibid.). Syntactically, a notable feature of *oxtšit* is the substitution of grammatical masculine forms with feminine forms (ibid.), which is the primary focus of the current study.

In Hebrew, there are two grammatical genders: masculine and feminine. Consequently, all nouns, whether animate or inanimate, possess grammatical gender and obligatorily maintain gender agreement with associated pronouns, adjectives, and verbs (Muchnik, 2016). This rule system poses a challenge for expressing fluid sexual and gender identities through language. However, LGBTQ+ Hebrew speakers have found creative ways to circumvent these grammatical constraints, enabling them to express their fluid gender identities while adhering to the rules of standard Hebrew (Bershtling, 2014).

Indeed, the primary syntactic feature of *oxtšit* is the ongoing alternation between masculine and feminine forms during conversation (Brom, 2022). While both Levon (2012) and Brom (2022) agree on the discursive contexts in which *oxtšit* is employed, such as telling jokes or talking about sex, there is a discrepancy between the two regarding the functions of the *oxtšit* variety. Levon (2012) claims that *oxtšit* serves to distance the speaker from the *oxtša* persona, as evidenced by speakers' use of *oxtšit* in the third person, i.e., when referring to others, particularly as a form of in-group mockery, but not in the first person, when referring to themselves, thereby reaffirming heteronormative masculinity. In contrast, Brom (2022) found *oxtšit* to be used in both first- and third-person, leading him to the conclusion that *oxtšit* is used not only to describe others (as argued by Levon), but also to embrace the *oxtša* persona, and thus to challenge heteronormative masculinity.

In the day and age of social media, the LGBTQ+ community faces a heightened risk of bullying and harassment (DeVito et al., 2018). The social image of gay men poses a threat to heterosexual men, leading to homophobic occurrences online and offline (DeVito et al., 2018; Kaplan & Offer, 2022).

Returning to the functional contrast between Levon (2012) and Brom (2022), if, as Levon (2012) contends, the use of *oxtšit* serves to distance individuals from the *oxtša* persona, such distancing might inadvertently exacerbate stigmatization within and of the LGBTQ+ community. Conversely, if, as argued by Brom (2022), individuals utilize *oxtšit* to embrace their *oxtša* identity, online communities where this slang variety is used could serve as an empowering platform for LGBTQ+ individuals to challenge heteronormativity, homophobia, and transphobia, while exploring and affirming their nonnormative identities. In this study, we center on grammatical gender alternations between masculine and feminine forms to unveil their functions, aiming to assess whether the investigated Facebook community provides a supportive and safe space for LGBTQ+ individuals or instead mirrors and perpetuates stigmatizing discourses and behaviors seen offline.

3. The Current Study

3.1 Data and Method

Following Androutsopoulos (2008), we conducted a *discourse-centered online ethnography* to scrutinize the functions of grammatical gender alternations online. In this paper, we present our analysis of user-generated content from a closed Facebook group with over 3000 members, primarily gay men, including bisexual and transgender men. Focused on the "bear community," a subculture within the gay community, the group promotes body positivity among gay men, especially embracing larger and hairier body types, and encourages open discussions on various topics while prohibiting disrespectful or violent interactions. With Author 1, a gay man and longstanding member of the group, we obtained permission from the group's administrator. To protect user anonymity, pseudonyms were assigned to group members, and all personal information was anonymized. After obtaining permissions and anonymizing the data, we collected about 200 posts and comments spanning six years that depict shifts in grammatical gender from masculine to feminine forms among group members. We present examples illustrating the functions of this linguistic phenomenon and address discrepancies between Levon (2012) and Brom (2022). Examples are presented in three lines: transliteration, gloss, and translation, following the Leipzig Glossing Rules (Max Planck Institute for Evolutionary Anthropology, 2015).

3.2 Findings

Our findings unveil a nuanced perspective on the utilization of grammatical gender alternations. While some group members employ the feminine form not only in the third person, when referring to others, but also in the first person, when referring to themselves, others engage with it differently. Consider the following example. A community member posted the following text:

(1) Example 1: Feminine forms for embracing the *oxtša* persona

imaaaaa, kama *še-ani* *išššššša!*
mother, how.much that-I woman!
'Mamaaaaa, I am such a wommmmman!'

This post garnered 18 likes and 18 comments. In the first comment, another member responded with a question:

ha-dub-a *ha-gdol-a?*
DEF-bear-F.SG DEF-big-F.SG?
'The big bear?'

The original poster then responded with five consecutive, single-word comments, each featuring a different adjective in the feminine form:

agal~gal-a
round~ATT-F.SG
'roundish'

av-a
thick-F.SG
'thick'

se'ir-a
hairy-F.SG
'hairy'

dub-i-t
bear-ADJ-F.SG
'bearish'

naš-i-t
woman-ADJ-F.SG
'feminine'

This repetition of grammatical feminine forms by the writer illustrates his identification with the *oxtša* persona and challenges heteronormativity. Remarkably, four out of the five adjectives used are not typically associated with the *oxtša* persona, which is often depicted as skinny and smooth. This underscores how members use the online community to defy societal stigmas and freely express their nonnormative identities.
However, our micro-analysis revealed that the usage of *oxtšit* in this group is more intricate than initially perceived. The following examples illustrate this complexity.

(2) Example 2: Alternations between feminine and masculine forms

The example comprises a storytelling post, divided into three paragraphs, recounting the writer's previous night. The three-paragraph-post is introduced with the following line in first-person feminine form:

ani me'ohev-et 😊
I in.love-F.SG 😊
'I'm in love 😊'

Subsequently, in the first paragraph, the writer uses third-person masculine form to describe a 'perfect bear' he had met:

etmol hikar-ti
yesterday meet.PST-1SG
dub-on mušlam
bear-DIM.M.SG perfect.M.SG
'Yesterday I met a perfect bear'

The writer goes on using 3rd person masculine form to enumerate the features he liked best about this bear:

hu haya yoter male […]
he be.PST.M.3SG more chubby.M.SG […]
'He was chubbier [than on his profile pictures and that was]'

afilu yoter seksi
even more sexy.M.SG
mi-ma še-tsipi-ti
than-what that-expect.PST-1SG
še-hu yihiye
that-he be.FUT.M.3SG
'even more sexy than what I had expected him to be'

yeš l-o xiyux maksim
POSS to-M.3SG smile charming.M.SG
še-yošev l-o al partsuf
that-sit.M.3SG to-M.3SG on face
agal~gal ve-mezukan […]
round~ATT.M.SG and-bearded.M.SG […]
'he has a charming smile that sits on his roundish and bearded face […]'

The second paragraph depicts how they went out to a pub and then spent the night together, detailing various erotic, physical acts described using masculine forms:

rakad-nu ve-nimrax-nu
dance.PST-1PL and-smear.PST-1PL
exad al ha-šeni […]
one.M.SG on DEF-other.M.SG […]
'we danced and were all over one another […]'

The writer then describes how 'at a certain point, when the sexual lust won over the desire to dance'

xazar-nu el-ay
return.PST-1PL to-1SG
ve-hizdayan-nu
and-have.sex.PST-1PL
ad or ha-boker
until light DEF-morning
'we went back to my place and had sex until dawn'

[…] nirdam-nu
[…] fall.asleep.PST-1PL
exad bi-yed-ei ha-šeni
one.M.SG in-arm-PL DEF-other.M.SG
ve-niš'ar-nu mekurbal-im
and-stay.PST-1PL cuddled-M.PL
ad še-kam-nu
until that-get.up.PST-1PL
'[and then] we fell asleep in each other's arms and stayed cuddled until we got up.'

In the third paragraph, the writer reflects on the morning after, contemplating the events of the previous night:

axarei kos kafe šel boker
after cup coffee of morning
'After morning coffee'

hu halax l-o
he.M.3SG walk.PST.M.3SG DATETH-M.3SG
ve-ani yošev-et po al ha-sapa
and-I sit-F.1SG here on DEF-sofa
mesupek-et
satisfied-F.SG
aval im ta'am šel od […]
but with taste of more […]
'he left and I'm sitting here on the sofa satisfied but with a taste for more […]'

Notably, in this final paragraph, the writer uses the feminine form while describing himself but employs the masculine form to depict his partner.
Examining the post holistically, it becomes evident that the user opts to express stereotypically feminine emotions, such as love and romance, in the feminine form, while adopting masculine form when discussing erotic, physical

activities – thus distancing both his partner and himself from the *oxtša* persona in erotic, sexual contexts. The next example further illustrates this finding:

(3) Example 3: Alternations between feminine and masculine forms

The example begins with a post featuring a GIF from the television series "Stranger Things," in which actor David Harbour is seen walking towards the refrigerator in his underwear, taking out and drinking milk from a carton. The GIF is introduced with the line, "The real reason for watching Stranger Things," clearly implying a sexual, erotic connotation.
This post garnered 41 likes and 26 comments. In these comments, when the group members refer to the actor, emphasizing his physical appearance, they use the third-person masculine form. However, when they refer to themselves and describe their own romantic feelings, they opt to use the first-person feminine form. For example, one of the members comments:

eize guf mušlam
which body perfect.M.SG
'What a perfect body'

And then another member adds:

hu madhim
he amazing.M.SG
'he's amazing'

ve-ani me'ohev-et b-o
and-I in.love-F.1SG in-M.3SG
'and I'm in love with him'

aval gam mexur-a
but also addicted-F.SG
'but also addicted'

It follows from this example that the group members use the first-person feminine form to embrace the *oxtša* persona when referring to their own romantic feelings but opt for the third-person masculine form to describe the object of their sexual desires.

4. Conclusions

When considering the disparity between Brom (2022) and Levon (2012) in relation to this study, Brom's (2022) findings resonate more closely with our own. In our study, we observed *oxtšit* being utilized in both the first and third person, suggesting that speakers do not distance themselves from the *oxtša* persona as Levon (2012) contended. Consequently, speakers employ the online platform to confront heteronormative masculinity at least to some extent.
However, it would be premature to discredit Levon's (2012) findings entirely. Although it is evident that speakers are no longer distancing themselves from the *oxtša* persona, our findings indicate that they still do not fully embrace it. More specifically, while the *oxtša* persona is used freely to describe stereotypically feminine emotions, such as love and romance, it is nevertheless excluded from descriptions of physical, erotic acts. Consequently, we argue that, to a certain degree, homophobia persists, even among first-person users of *oxtšit*. By perpetuating a stereotype of attractive masculinity through this behavior, they render the *oxtša* persona sexually unattractive and undesirable, thereby reinforcing heteronormative masculinity, and distancing the *oxtša* persona from their sexual desires in a partner.
Hence, the study indicates that the online sphere provides a platform for the gay community to enact acts of subversion, enabling individuals to diverge from traditional masculinity and embrace their feminine facets within a purportedly homophobia-free closed environment. However, this study reveals that stigmatization nevertheless penetrates beyond superficial layers. Upon closer examination, it becomes evident that heteronormativity persists even within closed online gay communities, reflecting societal norms, and mirroring the images seen in real life, as the most desirable man in the room is still, most likely, not an *oxtša*.

5. References

Androutsopoulos, J. (2008). Potentials and limitations of discourse-centred online ethnography. *Language@ internet*, 5(8).

Bershtling, O. (2014). "Speech creates a kind of commitment": Queering Hebrew. In L. Zimman, J. Davis, & J. Raclaw (Eds.), *Queer excursions: Retheorizing binaries in language, gender, and sexuality*. Oxford University Press, pp. 35--61.

boyd, d. (2011). Social network sites as networked publics: Affordances, dynamics, and implications. In Z. Papacharissi (Ed.), *A networked self: Identity, community, and culture on social network sites*. Routledge, pp. 39--58.

Brom, D. (2022). *Oxtšit* – A Queer Hebrew slang. *Israel Studies in Language and Society*, 15, pp. 160--180. [in Hebrew]

DeVito, M. A., Walker, A. M., & Birnholtz, J. (2018). "Too gay for Facebook": Presenting LGBTQ+ identity throughout the personal social media ecosystem. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), pp. 1--23.

Dori-Hacohen, G., & Shavit, N. (2013). The cultural meanings of Israeli tokbek (talk-back online commenting) and their relevance to the online democratic public sphere. *International Journal of Electronic Governance*, 6(4), pp. 361--379.

Duguay, S. (2016). "He has a way gayer Facebook than I do": Investigating sexual identity disclosure and context collapse on a social networking site. *New Media & Society*, 18(6), pp. 891--907.

Fox, J., & Ralston, R. (2016). Queer identity online: Informal learning and teaching experiences of LGBTQ individuals on social media. *Computers in Human Behavior*, 65, pp. 635--642.

Hall, K., & Nilep, C. (2015). Code-switching, identity, and globalization. In D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis*. John Wiley & Sons, pp. 597--619.

Harvey, K. (1998). Translating Camp talk: Gay identities

and cultural transfer. *The Translator*, 4(2), pp. 295--320.
Hiebert, A., & Kortes-Miller, K. (2023). Finding home in online community: Exploring TikTok as a support for gender and sexual minority youth throughout COVID-19. *Journal of LGBT Youth*, 20(4), pp. 800--817.
Kaplan, D., & Offer, S. (2022). Masculinity ideologies, sensitivity to masculinity threats, and fathers' involvement in housework and childcare among US employed fathers. *Psychology of Men & Masculinities*, pp. 399--411.
Levon, E. (2012). The voice of others: Identity, alterity and gender normativity among gay men in Israel. *Language in Society*, 41(2), pp. 187--211.
Marciano, A. (2010). The role of internet newsgroups in the coming-out process of gay male youth: An Israeli case study. In E. Dunkels, G. M. Franberg, & C. Hallgren (Eds.), *Youth culture and net culture: Online social practices*. IGI Global, pp. 222--241.
Max Planck Institute for Evolutionary Anthropology. (2015). *The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses*. Department of Linguistics, Leipzig. Retrieved June 13, 2024, from https://www.eva.mpg.de
Muchnik, M. (2016). Trying to change a gender-marked language: Classical versus Modern Hebrew. In J. Abbou & F. Beider (Eds.), *Language and the periphery*. John Benjamins, pp. 26--47.
Vider, S. (2013). "Oh Hell, May, why don't you people have a cookbook?": Camp humor and gay domesticity. *American Quarterly*, 65(4), pp. 877--904.

AI device for deradicalization process

Andrea Russo
Sorbonne University & CNRS
Pierre and Marie Curie Science department & GEMASS, 75005, Paris, France,
Andrea.russophd@gmail.com

Abstract

This collaborative effort highlights the innovations in our digital ethnography, where we pursued dual objectives: immersing in a Telegram radical community and deradicalizing an extremist group using AI technologies, specifically GPT-3.5. The methodology leverages classic chatbots, enhanced by advanced algorithms, to quantitatively measure the success of the deradicalization process. Integrating qualitative ethnographic observation and quantitative computational methodologies, my approach employs AI as virtual ethnographers to engage with the radical community. Notably successful, it stimulated debate and polarization, prompting internal discussions on potentially harmful actions. Consequently, some members exited, leading to the group's shutdown by the administrator.

Keywords: Sociology, Computational Methods, AI, Complex systems, Telegram.

1. Introduction

Radicalization, defined as the adoption of extreme ideologies condoning violence for political or ideological purposes, thrives in digital networks (Serafim, 2005). Online platforms like Telegram serve as hubs for the 'normalization' of inappropriate and illegal content, thanks to the ease of content sharing and dissemination as well as the sense of anonymity and 'safety-in-number' fostered by the platform (Semenzin and Bainotti, 2020a). Algorithms amplify this spread, enabling rapid dissemination and the creation of deceptive content like deepfakes. Despite abundant literature on digital radicalization, scant attention is paid to de-radicalization. This paper aims to fill this gap by presenting an action-research study utilizing AI to facilitate de-radicalization in digital environments.

Radicalization is a phased and complex process in which individuals or groups embrace a radical ideology condoning violence for specific political or ideological goals (Serafim, 2005). It can manifest at different levels: individual, involving identity issues, integration failures, and alienation; and governmental or political, seen in state-enforced or public opinion-driven radicalization (Extremists, 2024; Commission, 2024). This process is highly individualized, lacking a single pathway and taking various forms, influenced by a mix of triggers and drivers (Commission, 2024). The digital era has expanded opportunities for the dissemination of extreme content, with online platforms utilized for recruitment, psychological manipulation, and exposure to violent material (Blood, 2024; Area, 2024). Radicalization in digital environments is seen as a product of algorithmic culture, defined as "the use of computational processes to sort, classify, and hierarchize people, places, objects, and ideas, and also the habits of thought, conduct, and expression that arise in the relationships to those processes" (Hallinan and Striphas, 2014).

Studies also highlight the presence of algorithm or filter bubbles around conspiracy theories' videos and the algorithmic aggregation of problematic content (Faddoul et al., 2020; Matamoros-Fernández and Farkas, 2021; Ribeiro et al., 2020). The dissemination and consumption of radical content on social media are closely tied to smart devices, particularly social bots (Graham et al., 2020). Despite the frightening nature of this scenario, empirical research has significantly reduced its scale. Concerning YouTube's status as a 'radicalization and propaganda machine', scholars have shown otherwise. Many computational analyses revealed that users are predominantly exposed to mainstream news rather than extreme content (Hosseinmardi et al., 2021; Munger and Phillips, 2022). Additionally, major social media platforms actively moderate content to counter extremism, such as YouTube's counter-extremist project. Regarding the impact of bots on disinformation processes, scholars emphasize the need for deeper empirical investigations (Krizhevsky et al., 2012). Qualitative methods are crucial for understanding people's reactions to bots and the misinformation they propagate (Caliandro et al., 2024).

However, the empirical account largely applies to public social media environments accessible to civic groups, academics, and journalists. Enclosed and invisible spaces like private WhatsApp or Telegram groups present different challenges. Telegram, for instance, is noted for hosting various controversial groups (Rogers, 2020; Semenzin and Bainotti, 2020b), where hate speech and propaganda are widespread (Vergani et al., 2022). Therefore, understanding radicalization in digital environments requires considering it as a cultural process alongside its algorithmic aspects. Deradicalization is a process where individuals renounce previously embraced ideologies. It's crucial to consider group membership and the inter-group context forming radicalization's basis. Radical groups typically exhibit a strong in-group identity and perceive an out-group as responsible for in-group grievances, legitimizing violent attacks for societal and political changes (Doosje et al., 2016). Importantly, individuals resist radical ideologies to the extent of their resilience. There's scant literature on initiating deradicalization with online radical groups, especially leveraging digital and computational methods. My research aims to fill this gap.

2. Data & Methodology

After an initial observation that confirmed the inherent violence of the group, I decided to collect data through the

use of Telegram's API. The full dataset is available in Zenodo repository (Russo, 2024b). The first phase is obviously to observe and test whether the group is radicalised. After some time, elements of the group showed extreme behaviour, evidencing a desire to even make violent and dangerous gestures. After ascertaining that the group was radicalised and violent, I began with data collection.

I start to gathered data from the Italian chat group 'Put down the covid-mask' on Telegram.

The group *'Put down the covid-mask'* has about 450 people (2022/11/28) with an average of about 70 people online daily. From 2022/10/14 until 2022/11/02, I collected data to assess the initial conditions of the social system. Then I started to observing the group from 2022/12/01 for obtaining valuable data for the deradicalisation strategy and GPT training, while the active deradicalisation phase spans from 2023/3/23 to 2023/5/07. I Divided tasks into periods as show in Table 1:

Regarding the activity of my bots/profiles, during my interaction phase, I aimed to gain social status by presenting the characteristics of my bots and answering questions clearly. For example, I would say, 'As a doctor, I don't think this is correct.' If the conversation continued, I responded with well-structured sentences to assert my position. While, during the interaction phase, I adapted GPT's sentences to fit the context and situation. In the final stage, my role was to control GPT's responses.

Sentiment analysis, already used to calculate hate speech in different work (Del Valle, 2023; Subramanian et al., 2023), was performed on the text using Italian VADER. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a tool used for sentiment analysis of text. It is specifically designed to analyze sentiment in social media content by assigning a sentiment score to text based on a lexicon of words with predefined sentiment scores. VADER can determine if text expresses positive, negative, or neutral sentiment, along with the intensity of each sentiment, and it was originally developed with an English lexicon, but there have been efforts to adapt it to other languages, including Italian.

An Italian lexicon has been curated by my self (Russo, 2024a) specifically for sentiment analysis using VADER, allowing it to analyze sentiment in Italian text more accurately.

Regarding the deradicalization process, I or this work not intended to manipulate people's thinking, but rather to foster a socialization dynamic that encourages people to discard dangerous and harmful values. Manipulation uses deceit or coercion to change someone's perspective, prioritizing the manipulator's interests over others, and it focuses on control, lacking transparency and respect. (Simon and Foley, 2011; Braiker and others, 2004) In contrast, socialization involves open, respectful engagement to share ideas and values. It aims to build mutual understanding and foster genuine change through honest communication and empathy. (John, 2001) Neither my responses nor those from GPT were intended to coercively share a message or make it appealing, provoking a manipulation effects (Van Dijk, 2006). Instead, my goal was to spur discussion on the topic by presenting different perspectives, encouraging people to reflect on their actions or opinions.

For example, the responses have never been *"don't do this because it is wrong"* but rather *"it is not recommended to do this because it can have negative effects on you or your loved ones"*. [1] In this work, the information shared within the group aims not to persuade, but to present diverse perspectives. By doing so, it offers participants the opportunity to reconsider and potentially transform their views on various topics.

Regarding the strategy of deradicalization, I used the Juan Pujol García fake-agent network Intelligence-method (Domenico, 2022). He crafted fake agents who produced deceptive reports on Britain from diverse sources to Nazis. (Domenico, 2022). In emulation of GARBO's strategy, I employ controlled fictitious accounts to build a network. I use five accounts with the same strategic principle. To enhance credibility, each account is endowed with a story crafted from information gathered through qualitative sociological analysis and the anthropological "follow the native" principle. The "follow the native" priciple is a Caliandro et al.'s 2018 work (Caliandro, 2018) that highlights how embracing the principles of "follow the medium" and "follow the natives" offers valuable strategies for ethnographers navigating social media environments. "Follow the medium" involves leveraging the Internet's natural logic, like tags and hashtags, for data gathering and analysis. Conversely, "follow the natives" entails observing how social actors construct the social order. These principles yield two key strategies:

1. Observing and describing online communication processes structured by social media affordances and digital devices, aligned with "follow the medium".

2. Understanding online social formations arising from diverse digital device practices, along with the meanings users attribute to activities within these formations, in line with "follow the natives".

I opted for the GPT-3.5 Davinci model for my AI system, connected with Telegram API, aiming to improve conversational agents in bot-human interactions, providing more natural and engaging conversations that better meet user needs (Bender et al., 2021; Brown et al., 2020). The linguistic aspect is significant since ChatGPT generates responses based on probabilistic information without explicit reference to meaning, resembling a "stochastic parrot" (Bender et al., 2021).

3. Results

To observe the phase transition in the system, I collected data in both the initial condition of the system and the phase after the deradicalization process.

3.1. The initial condition

Analyzing data from 2022/10/14, around 1,300 interactions revealed initial system conditions—reference network, prevalent words, topics, ecosystem, and influence

[1] Example of a response actually used to discourage possible violent actions, such as setting hospitals on fire.

Table 1: Planned activities and their durations

Time	Activity	Duration (days)
2022/12/01 to 2023/3/22	Observing task	112
2023/3/23 to 2023/4/10	I interact alone task	18
2023/4/10 to 2023/4/20	I interact with GPT suggestion task	10
2023/4/20 to 2023/5/07	GPT interaction only task	17

network. Figure 1 illustrates the reply network, revealing a central hub identified as the group administrator and reveals group influencers (a bigger node exert more influence due to more connection and interaction with others nodes) and associated communities (communities are defined as the tendency to interact toward a particular group of people than other nodes outside the group). Notably, a majority of comments are succinct, while a few individuals contribute extensive text (the hub), as express in violet color-community. This observation offers valuable insights into the group's structural composition and community interactions ,given that there is a hub disproportionate to the other nodes.

Regarding message content, I assessed the group's overall sentiment, the unusual value is the neutral sentiment (markedly very low), while there is a notably high negative and positive value (the latter is achieved by means of ironic phrases that VADER defines as positive). There is a substantial presence of comments endorsing weapons, revolt, and conspiracy theories such as those involving George Soros and Klaus Schwab such as the Global Control, Population Microchipping and "The Great Reset".

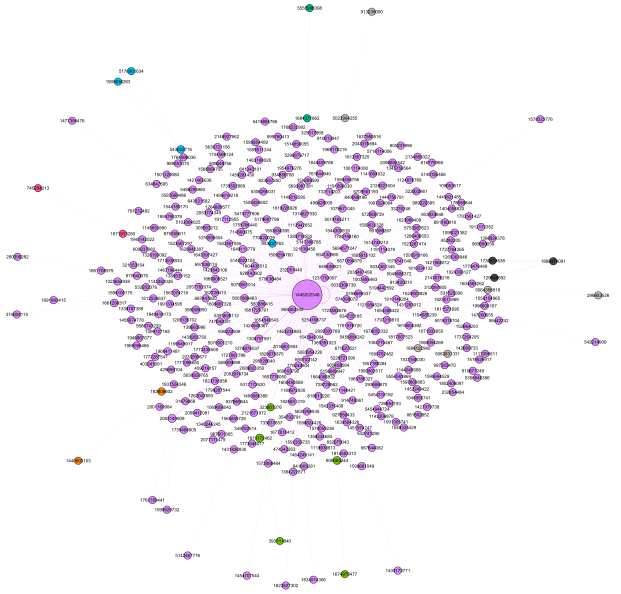


Figure 1: Network account from *"Put down the covid-mask"*

3.2. After Dynamics

The initial conditions of the system and the observed social dynamics unfortunately described a difficult and hostile environment for my goals. I admit I was discouraged by the possible results from after the interaction with my bots and ChatGPT, but starting with the channel network after the deradicalisation dynamics, as shown in Figure 2 there was a change from the initial conditions.

While the administrator remains the main hub of the network, their centrality has decreased significantly as show in Table 2. This Table, indicates the variation in weight[2] and importance in the network during the two phases. Since the implementation of my accounts and the dynamics they bring, the relative weight of other accounts in the network has shifted considerably. As shown in Figure 2, the grey nodes representing my accounts indicate that interactions with other accounts cover roughly $\sim 17\%$ of the entire community (represented by light green and light blue nodes). Concerning accounts displaying violent behavior, I encountered several, with two particularly prominent ones represented by dark orange nodes. Despite their violent tendencies, frequent interactions occurred, forming a cluster in the graph's top center where open discussions took place. This led to the formation of new, smaller communities (light green, light blue, and dark green nodes) distinct from the violet one directly connected to the administrator—an 'anti-

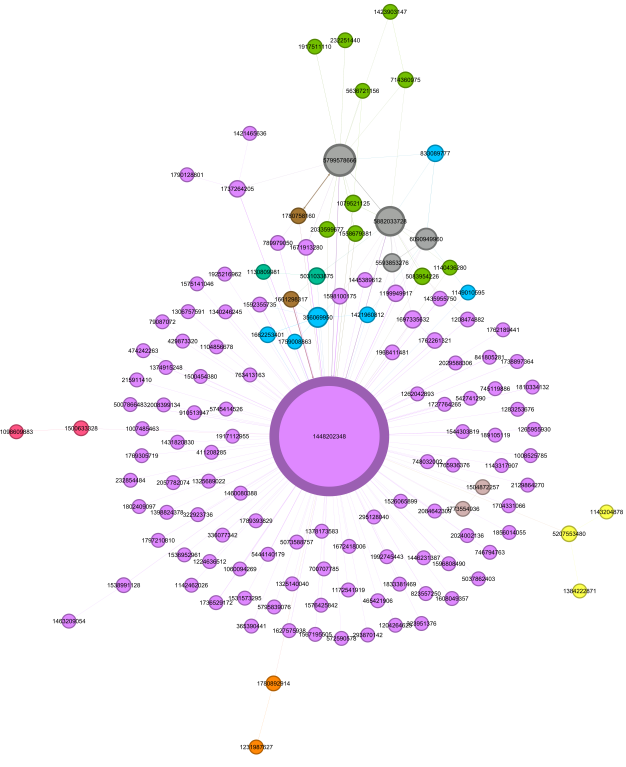


Figure 2: After dynamics accounts with violent reactions and behaviour

[2]Based on the number of edge for a node, but ponderated by the weigtht of each edge.

Table 2: Degree and Weighted degree before and after dynamics

Accounts	D. before dynamic	D. After dynamic	After D. Weighted degree
Administrator	349	122	222
My account	-	22	68
My account	-	19	39
My account	-	10	14
Admin's likely collaborator	9	8	8
My account	-	6	13
Acc. with Violent behaviour	-	3	33
Acc. with Violent behaviour	3	3	16

Table 3: Average text sentiment of the Telegram group before and after dynamics

Timing	Negative	Neutral	Positive	Compound
Initial condition	0,125332436	0,724032301	0,149318977	0,481011844
After dynamics	0,078551515	0,938115152	0,104549495	0,067783434
Variation	-0,046780921	0,214082851	-0,044769482	-0,41322841

echo chamber.

Significant results emerged in sentiment analysis. Table 3 illustrates the variations between the initial condition of the system and the post-dynamic sentiment outcomes. The findings indicate a decline in negative sentiment, an uptick in neutral sentiment, and a corresponding reduction in positive sentiment. Additionally, the compound sentiment appeared more negative post-dynamic, attributable to increased neutral sentiment fostering balanced and less radical discussions within the group.

4. Conclusion

This paper provides valuable insights into the setup of various strategies, particularly deradicalisation strategies, by examining the initial conditions of the social network system. Given the scarcity of publications on deradicalisation processes in online platforms, this work stands out as innovative and relevant in today's context.

The integration of AI technologies offers significant advantages in implementing strategic approaches, providing practical insights into the interplay of social dynamics, artificial intelligence, and non-linear dynamics, with the possibility to replicate the study in different contexts with the same methodology ad tools. The complex dynamics presented in this paper highlight the need for further analysis, considering external events' impact on group dynamics, emphasizing the importance of technical feasibility and understanding unpredictable dynamics for societal and institutional security,

In conclusion, this work contributes significantly in:

1. Improving interaction strategies through understanding values and social dynamics;

2. Gaining insights for deradicalization by analyzing community and network structures;

3. Investigating AI's role in the deradicalization process within social networks.

5. Copyrights & Acknowledgements

Proceedings and data will be published under a Creative Commons Attribution 4.0 International license. This project was fully funded by the University of Catania (Italy).

6. References

Area, R. (2024). Radicalisation in the digital era | rand.

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623.

Blood. (2024). Radicalisation and extremism | cambridgeshire and peterborough safeguarding partnership board.

Braiker, H. B. et al. (2004). Who's pulling your strings?: How to break the cycle of manipulation and regain control of your life. Technical report, McGraw-Hill.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Caliandro, A., Gandini, A., Bainotti, L., and Anselmi, G. (2024). The platformisation of consumer culture: A digital methods guide. In *The Platformisation of Consumer Culture*. Amsterdam University Press.

Caliandro, A. (2018). Digital methods for ethnography: Analytical concepts for ethnographers exploring social media environments. *Journal of contemporary ethnography*, 47(5):551–578.

Commission, E. (2024). Prevention of radicalisation - european commission.

Del Valle, E. e. a. (2023). Sentiment analysis methods for politics and hate speech contents in spanish language: a systematic review. *IEEE Latin America Transactions*, 21(3):408–418.

Domenico, V. (2022). Garbo, la spia che rese possibile lo

sbarco in normandia - sistema di informazione per la sicurezza della repubblica.

Doosje, B., Moghaddam, F. M., Kruglanski, A. W., De Wolf, A., Mann, L., and Feddes, A. R. (2016). Terrorism, radicalization and de-radicalization. *Current Opinion in Psychology*, 11:79–84.

Extremists, V. (2024). Prevention of radicalisation - european commission.

Faddoul, M., Chaslot, G., and Farid, H. (2020). A longitudinal analysis of youtube's promotion of conspiracy videos. *arXiv preprint arXiv:2003.03318*.

Graham, T., Bruns, A., Zhu, G., and Campbell, R. (2020). Like a virus: The coordinated spread of coronavirus disinformation.

Hallinan and Striphas. (2014). Algorithmic culture, defined as "the use of computational processes to sort, classify, and hierarchize people, places, objects, and ideas, and also the habits of thought, conduct, and expression that arise in the relationships to those processes".

Hosseinmardi, H., Ghasemian, A., Clauset, A., Mobius, M., Rothschild, D. M., and Watts, D. J. (2021). Examining the consumption of radical content on youtube. *Proceedings of the National Academy of Sciences*, 118(32):e2101967118.

John, M. (2001). *Sociology*. Brossura.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105.

Matamoros-Fernández, A. and Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & new media*, 22(2):205–224.

Munger, K. and Phillips, J. (2022). Right-wing youtube: A supply and demand perspective. *The International Journal of Press/Politics*, 27(1):186–219.

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., and Meira Jr, W. (2020). Auditing radicalization pathways on youtube. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 131–141.

Rogers, R. (2020). Deplatforming: Following extreme internet celebrities to telegram and alternative social media. *European Journal of Communication*, 35(3):213–229.

Russo. (2024a). Github - andrearussoagid/vader-italian-sentiment: Under the [mit license] in this vader version, i have created an italian vader thanks to some modifications to the code and the addition of a special italian dictionary.

Russo, A. (2024b). Telegram_Italian_NoVax: AI-Influencer to mitigate radicalism and social threats on Telegram. *Zenodo*.

Semenzin and Bainotti. (2020a). Online platforms like telegram serve as hubs for hate speech and extremism, exploiting features like content sharing and anonymity.

Semenzin, S. and Bainotti, L. (2020b). The use of telegram for non-consensual dissemination of intimate images: Gendered affordances and the construction of masculinities. *Social Media+ Society*, 6(4):2056305120984453.

Serafim. (2005). Radicalization, defined as the adoption of extreme ideologies condoning violence for political or ideological purposes.

Simon, G. K. and Foley, K. (2011). *In sheep's clothing: Understanding and dealing with manipulative people*. Tantor Media, Incorporated.

Subramanian, M., Sathiskumar, V. E., Deepalakshmi, G., Cho, J., and Manikandan, G. (2023). A survey on hate speech detection and sentiment analysis using machine learning and deep learning models. *Alexandria Engineering Journal*, 80:110–121.

Van Dijk, T. A. (2006). Discourse and manipulation. *Discourse & society*, 17(3):359–383.

Vergani, M., Martinez Arranz, A., Scrivens, R., and Orellana, L. (2022). Hate speech in a telegram conspiracy channel during the first year of the covid-19 pandemic. *Social Media+ Society*, 8(4):20563051221138758.

Spoken vs. Written Computer-Mediated Communication

Hannah J. Seemann[1], Sara Shahmohammadi[2], Manfred Stede[2], Tatjana Scheffler[1]
[1] Department for German Language and Literature, Ruhr-University Bochum, Germany
[2] Department of Linguistics, University of Potsdam, Germany
hannah.seemann@rub.de shahmohammadi@uni-potsdam.de
stede@uni-potsdam.de tatjana.scheffler@rub.de

Abstract

Using PARADISE, a German corpus of thematically parallel blog posts and podcasts annotated for discourse features, we study discourse-level differences in spoken and written computer-mediated communication. We show that discourse relations as defined in models of discourse coherence such as Rhetorical Structure Theory as well as discourse particles can indicate the medium a text is taken from (podcast vs. blog post), and therefore the document's mode (spoken vs. written).

Keywords: speech vs. writing, computer-mediated communication, discourse structure, statistical analysis

1. Introduction

Since the description of orality as a continuum by Koch and Oesterreicher (1985) and others, many studies have been conducted on features that distinguish spoken and written communication. Such studies have identified several features that vary between speech and writing, and between different communicative settings (Biber et al., 1999; Kunz et al., 2018) – though most of these studies focus on lexical and situational features. With the emergence of digital communication (or computer-mediated communication, CMC), the focus of research interest has shifted to features of spoken language found in written forms of communication. This leaves two areas in the intersection of CMC and language mode (speech vs. writing) underexplored: (i) features specific to computer-mediated communication realized in the spoken mode, such as in spoken messages or podcasts; and (ii) linguistic differences between speech and writing that relate to the discourse level of communication.

If spoken CMC is studied, it is mostly in the context of multimodal communication, for example in work on the interaction between live streamers and the accompanying chat, or YouTubers reacting to previous comments, etc. (Heyd, 2021). Other forms of spoken CMC might seem at first rather similar to forms of oral discourse like taped interviews as described by the notion of *secondary orality*, introduced by Ong (1982). Secondary orality is different from traditional orality, as it is based on and linked to the practices of a group that is also used to written communication, meaning that (the sense) of orality can be used strategically and is not just the default mode of communication.

However, first register studies on podcasts have shown that they are different both from other forms of CMC as well as other forms of spoken media like broadcasts or conversations. They are different from most other forms of CMC as they are spoken, and they are different from other forms of spoken language as they combine characteristics of informal and spontaneous speech with narrative elements (Babyode et al., 2023). In addition, a study by Ortmann and Dipper (2019) on registers on the orality continuum (but not including podcasts) showed that even discourse-related features like co-reference can be used to distinguish between different modes of language.

Our analysis goes beyond the results by Ortmann and Dipper (2019), focusing on additional features that are related to the discourse level of a text. We fit a mixed-effects regression model to study how well discourse-related features such as the discourse structure of a text (in the form of discourse relations), as well as connectives and discourse particles (lexical features related to discourse) can predict the mode of a given document. As we find statistically significant effects for five out of eight groups of discourse relations and discourse particles, we conclude that distinguishing a document's mode based on discourse-related features is possible.

2. Data and Annotation

2.1. Corpus

To analyze the differences between spoken and written communication in the context of CMC, we use PARADISE (PARAllel DIScoursE) – a corpus of parallel blog posts and podcasts in German.[1] The corpus consists of 69 podcast transcripts and 69 corresponding blog posts from eight different data sources in the 'business' and 'science & culture' domains. The blog posts are written to introduce what is talked about in the podcast episodes, so the same types of content are covered in both media. Topics in the business domain range from digitalization to health and the food industry. These podcasts are scripted interviews rather than free conversations, with a host or group of hosts talking to a company internal or external guest who is an expert on the episode's topic, usually about 20 minutes per episode. The science & culture domain covers topics from various fields of academic research, astronomy in theory and practice, and German media and politics. The podcasts are either free 1:1 interviews by a host with an expert on the episode's topic or group conversations with varying members from a fixed pool of participants. Podcasts in this domain are between one and four hours in length.

We first manually determine the level of parallelism between blog posts and podcast transcripts on a per-sentence

[1] We make the corpus freely available at:
https://osf.io/59acq/

basis. For each discourse segment in a blog post, the corresponding section in the transcribed podcast is manually identified by one annotator and labeled for the type of parallelity by two annotators (from *category A: topic and wording identical* to *category C: blog and podcast share the topic, but one goes into much more detail* and *D: blog and podcast address roughly the same topic, but with different content*). We used weighted Krippendorff's α to calculate inter-annotator agreement (α = 0.53).[2] A third annotator assigned labels if both annotators did not agree on one label.
The distributions of parallelity labels in the corpus is shown in Table 1.[3]

A	B	C	D	E
6.87	58.47	27.51	6.38	0.73

Table 1: Distribution of parallelity labels, in %.

For our analysis, we focus on a random sample of 70 segments (about 33%) labeled with *category B: blog and podcast address the same topic, but one goes into more detail*, as this allows us to study how the same content is talked about in different ways in the two different modes of communication. The size of the resulting sub-corpus is shown in Table 2.

	Blogs	Transcripts	Total
Science	2,411	30,416	32,827
Business	788	4,814	5,602
Total	3,199	35,230	38,838

Table 2: Token count in our corpus, split by medium and domain.

The individual blog posts and podcast transcripts vary in their length. The mean length for blog posts is 70 tokens with a standard deviation (SD) of 44 tokens, the mean length for podcast transcripts is 756 tokens with a SD of 649 tokens.

2.2. Annotation

In this sub-corpus of parallel segments, we manually annotate the discourse features that are to be tested as predictors of mode.

Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) is a model of discourse structure that aims at capturing the coherence structure of a given text in the form of a hierarchical tree of discourse relations that span between segments of this text. At the lowest level, discourse relations hold between so-called *Elementary Discourse Units* (EDUs), often clauses.

(1) [I just had to eat the entire cake] [because it was so delicious!]

In example (1), the two EDUs are indicated by the brackets. The relation holding between them is a CAUSAL relation, the second EDU presents a state of affairs causing the event described in the first EDU. Identifying the discourse relations makes it possible to describe the structure of a text, from which a reader might gain insights into the intentions of its author. How many different relations there are differs between the annotation frameworks, though most frameworks agree on the types of relations applying to a text. In our annotation, we follow the guidelines by Stede (2016). Given that RST is mainly designed for written text, we expand the annotation schema to include a COMPLETION relation that spans between segments of an interrupted utterance, allowing us to account for some particularities of spoken language. For our analysis, we group the relations by their semantic category, as shown in Table 3.

Semantic Category	**Discourse Relations**
CAUSAL	Cause, Justify, Evidence, Reason-N, Reason, Result
CONTRAST	Antithesis, Concession, Contrast
HYPOTHETICAL	Condition, Enablement, Means, Motivation, Otherwise, Purpose
JUDGEMENT	Evaluation-N, Evaluation-S, Interpretation, Solutionhood
TEMPORAL	Circumstance, Sequence
INFORMATION	Background, E-Elaboration, Elaboration, Preparation, Restatement, Summary Question
ADDITIVE	Conjunction, Joint, List
STRUCTURAL	Attribution, Completion, Sameunit

Table 3: Discourse relations annotated in our corpus, grouped by semantic category.

Figure 1 shows the proportions of discourse relations in both modes. E.g., of all relations annotated in a written document, on average 31% of them are in the semantic category of INFORMATION relations. Overall, the proportion of a certain relation group is generally lower in our spoken documents, meaning that we find a broader variety of relations in spoken than in written documents. However, this distribution is not necessarily an effect of a document's mode and more likely to stem from the spoken documents typically being longer than the written ones.
The annotation of discourse relations was conducted by one annotator and a second annotator reannotated parts of the corpus to evaluate the quality of the annotations. To calculate the agreement between the two annotators, we used RST-Tace (Wan et al., 2019). With $\kappa = 0.49$, our inter-

[2]However, the two annotators only disagreed substantially in 33 out of 406 segments (8.12%). In 150 cases (36.94%), the annotators assigned neighboring labels.

[3]*Category E: not parallel* is applied in the few cases previously considered as potentially parallel that do not fit the labels A–D.

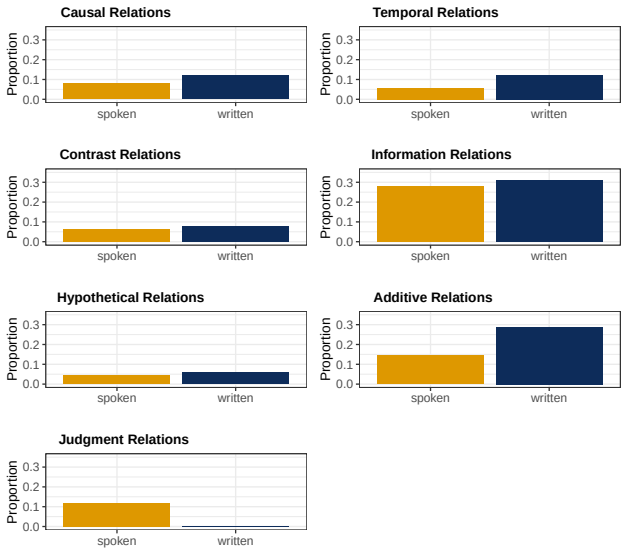


Figure 1: Distribution of discourse relations, by mode. Please note that the proportions do not add up to 1 due to STRUCTURAL relations being excluded from this graphic.

annotator agreement is comparable to similar complex annotation tasks. (A direct comparison to other annotations of RST-style coherence relations is difficult, either because no agreement is reported or a different evaluation method is used.)

Besides the representation of discourse coherence in the form of RST relations, we annotated two lexical features typically associated with discourse structure.

Discourse connectives are lexical items that signal the type of discourse relation holding between segments. They can belong to various syntactic categories such as conjunctions, adverbs, or prepositional phrases. In example (1), the connective *because* explicitly marks the relation to be CAUSAL. Discourse relations can be signaled by many different connectives, and one connective usually can signal more than one relation type. Still, as every connective signals at least one relation and the proportion of explicitly marked relations differs between modes (Tonelli et al., 2010), the frequency of connectives might help in distinguishing between spoken and written language.

We base our annotation of connectives on DiMLex, the lexicon of German discourse markers (Stede, 2002; Scheffler and Stede, 2016). In all of our texts, we annotated 1,117 instances of connectives in total, with *und* ('and'), *aber* ('but'), and *wenn* ('if') being the most frequent connectives. For our analysis, we use only the total frequency of connectives relative to the total token count of the respective document. While the kinds of connectives used may also shed light on the mode differences, we do not use this information here, because (i) it partially overlaps with the RST discourse relation types, (ii) many connectives are very rare yielding sparse data problems, and (iii) we want to focus on non-lexical, structural features for distinguishing speech and writing.

Discourse particles[4] in German are sentence modifiers that are used for Common Ground management (Stalnaker, 2002) and to signal the epistemic states of the interlocutors (Zimmermann, 2011).

(2) Ich muss noch einmal backen, jemand hat **ja** den
I must once again bake someone has JA the
ganzen Kuchen von gestern aufgegessen.
entire cake from yesterday eaten.up
'I have to bake once again because someone ate the entire cake made yesterday (and we all know who ate the cake).'

(3) Ich muss noch einmal backen, jemand hat **wohl**
I must once again bake someone has WOHL
den ganzen Kuchen von gestern aufgegessen.
the entire cake from yesterday eaten.up
'I have to bake once again, it seems someone ate the entire cake made yesterday.'

In (2), the discourse particle *ja* is used to indicate that everybody involved in the conversation knows what is being talked about, whereas *wohl* in example (3) is employed to signal that the speaker is not entirely certain about what happened or just found out about it. Due to these functions, discourse particles have previously been argued to be associated with discourse structure (Karagjosova, 2004; Döring, 2016).

We annotated 24 different discourse particles[5] in our corpus, for a total of 518 instances, with *ja, eben* and *halt* being the three most frequent particles. Similarly to the connectives, we use the frequency of all particles relative to the total token count of a given document as a feature for our analysis. Because discourse particles are typically more frequently found in spoken conversation than in written texts, we expect the frequency of particles employed to be an indicator of sentence mode.

3. Analysis and Results

Using the R package *lme4* (Bates et al., 2015), we fit a mixed-effects regression model to test how well the discourse-level features described above are able to predict the mode (spoken podcast or written blog post) of a given text. Discourse relations are included as proportional measures to represent the distribution of relations over a given document. As the category of structural discourse relations – ATTRIBUTION, COMPLETION, SAMEUNIT, which are added for text-specific reasons – differs from the other semantically and text-immanently motivated relation groups, we do not include this category in our analysis. In a second step, we test whether including the lexical discourse features discourse connectives and particles improves the fit of the model.

We include the document's domain (= business/science) as random factor, allowing for varying intercepts for the groups of data sources. Due to the relatively low number

[4] Also called *modal particles* or *Abtönungspartikeln* in German (Engl. 'shading particles').

[5] *aber, allerdings, auch, denn, doch, eben, eh, eigentlich, einfach, gleich, halt, irgendwie, ja, jetzt, leider, mal, schon, selbst, sogar, tatsächlich, vielleicht, wahrscheinlich, wirklich, wohl*

of observations, we do not include random slopes in our model. All frequencies are z-standardized to account for the different scales of our predictors.
Our baseline model with the seven groups of discourse relations as predictors for mode is shown in Table 4.

Fixed Effects	Est.	Std. Err.	z-val.	p-val.
(Intercept)	1.066	0.858	1.242	0.214
Causal	-3.512	1.312	-2.676	0.007**
Contrast	-2.784	0.979	-2.841	0.004**
Hypothetical	-2.024	0.997	-2.030	0.042*
Judgement	0.987	1.259	0.784	0.432
Temporal	-1.714	0.740	-2.316	0.020*
Information	-4.585	1.484	-3.089	0.002**
Additive	-4.796	1.514	-3.167	0.001**

Table 4: Results of the mixed-effects regression model: MODE $\sim$ C(DISCOURSE RELATIONS) + (1 | DOMAIN)

Adding discourse particles as a predictor did significantly improve model fit ($p < 0.005$), whereas adding connectives did not ($p = 0.38$). This means that while we find main effects for discourse relations and discourse particles, there is no main effect for connectives.
As the estimates for the single relation groups vary, Figure 2 provides an overview of each group's estimates and confidence intervals.

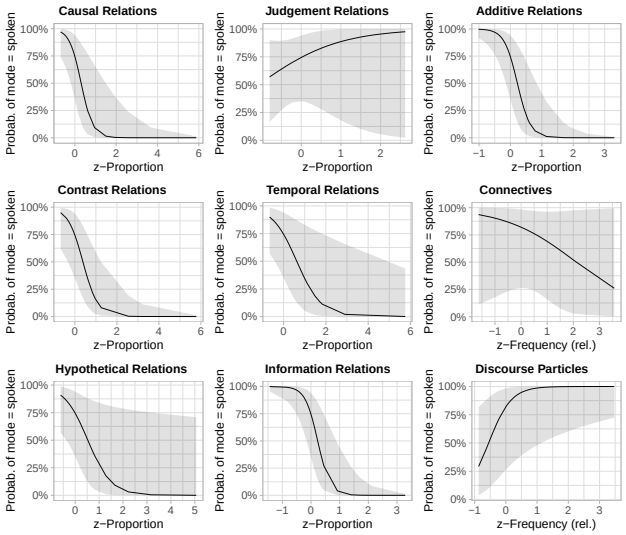


Figure 2: Probability of the mode being 'spoken' as predicted by the regression model, given the proportion of the particular relation for discourse relation and given the relative frequency for discourse particles and connectives. Proportion and frequency values are z-standardized.

4. Discussion

The model output suggests two trends: The predicted probability of the mode being 'spoken' is lower if the relation groups CAUSAL, CONTRAST, HYPOTHETICAL, TEMPORAL, INFORMATION, and ADDITIVE are more often present in the input data, and if there are more discourse connectives. The probability of mode = 'spoken' is higher if there are more JUDGEMENT relations and discourse particles. Except for JUDGEMENT relations and connectives, our mixed-effects regression model predicts these tendencies to be statistically significant. The regression lines in Figure 2 show that except for ADDITIVE, INFORMATION, and JUDGEMENT, all relation groups are relatively infrequent in the spoken podcasts. Given the register of our data, this is not surprising: As the corpus comprises podcasts that are conversations rather than discussions, it is fitting that we find fewer argumentative relations and more relation groups that contain informative or explicitly evaluative discourse relations. The blogs on the other hand might use more argumentative language to convince a possible reader to spend more time with the promoted content and to listen to the podcast – or the blog post might highlight the main point talked about in the podcast episode, which shifts the proportion of relations in favor of the more argumentative than informative relations.

For discourse particles, the results of our analysis match the prediction of the particles being suitable for distinguishing mode, as they are usually described as a phenomenon of spoken language, which is confirmed by our data. The connectives on the other hand do not seem to help in distinguishing the mode, at least not in addition to the discourse relation information. It should be noted though that the frequency of connectives used is not equal to the proportion of explicit to implicit relations, and studying this proportion might yield different results.

Our initial study focused on the variance in discourse-level features between spoken and written modes, though other factors have been shown to influence linguistic choice, too, such as register or extra-linguistic affordances (Biber and Conrad, 2019; Scheffler et al., 2022). Our results indicate that studying register variation should not be excluded when comparing language use on different positions of the orality continuum. Further studies on this corpus and larger samples of different registers are needed to delineate the exact distribution of discourse relations over various text types – and to study to what extent the coherence structure of a text is influenced by such factors. In addition to these open questions and due to the small sample size, which also leads to large confidence intervals in our model as seen in Figure 2, these results should be treated as interim results. However, our results nonetheless match previous results on other discourse-level features and show that distinguishing between two text modes is possible based on discourse-level features only. Including discourse structure in further studies of mode and register seems to be promising.

5. Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 317633480 – SFB 1287.

We would like to thank Elisa Lübbers and Daniel Foppe for their support in the annotation tasks and our anonymous reviewers for their detailed and helpful comments.

6. References

Babyode, A., Bosman, L., Chan, N., Ehret, K., Fong, I., Harris, N., Hewton, A., Reid, D., Taboada, M., and Wong, R. (2023). Structural linguistic characteristics of podcasts as an emerging register of computer-mediated communication. In *Proceedings of the 10th International Conference on CMC and Social Media Corpora for the Humanities*, pages 3–6, Mannheim, Germany.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Biber, D. and Conrad, S. (2019). *Register, Genre, and Style*. Cambridge Textbooks in Linguistics. Cambridge University Press, Cambridge, 2 edition.

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *The Longman grammar of spoken and written English*. Longman, London.

Döring, S. (2016). *Modal Particles, Discourse Structure and Common Ground Management*. Dissertation, Humboldt-Universität zu Berlin.

Heyd, T. (2021). Tertiary Orality? *Anglistik*, 32(2):131–147.

Karagjosova, E. (2004). *The Meaning and Function of German Modal Particles*. Dissertation, Universiteit Utrecht.

Koch, P. and Oesterreicher, W. (1985). Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch*, 36:15–43.

Kunz, K., Lapshinova-Koltunski, E., Martínez, J. M. M., Menzel, K., and Steiner, E. (2018). Shallow features as indicators of English-German contrasts in lexical cohesion. *Languages in Contrast: International Journal for Contrastive Linguistics*, 18(2):175–206.

Mann, W. C. and Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281.

Ong, W. (1982). *Orality and Literacy: The Technologizing of the Word*. Routledge, London.

Ortmann, K. and Dipper, S. (2019). Variation between different discourse types: Literate vs. oral. In Marcos Zampieri, et al., editors, *Proceedings of the Sixth Workshop on NLP for Similar Languages, Varieties and Dialects*, pages 64–79, Ann Arbor, Michigan. Association for Computational Linguistics.

Scheffler, T. and Stede, M. (2016). Adding semantic relations to a large-coverage connective lexicon of German. In Nicoletta Calzolari, et al., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1008–1013, Portorož, Slovenia. European Language Resources Association (ELRA).

Scheffler, T., Kern, L.-A., and Seemann, H. (2022). The medium is not the message: Individual level register variation in blogs vs. tweets. *Register Studies*, 4(2):171–201.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, 25:701–721.

Stede, M. (2002). DiMLex: A Lexical Approach to Discourse Markers. In Alessandro Lenci et al., editors, *Exploring the Lexicon - Theory and Computation*. Edizioni dell'Orso, Alessandria.

Stede, M. (2016). *Handbuch Textannotation: Potsdamer Kommentarkorpus 2.0*. Number 8 in Potsdam cognitive science series. Universitätsverlag Potsdam, Potsdam.

Tonelli, S., Riccardi, G., Prasad, R., and Joshi, A. (2010). Annotation of discourse relations for conversational spoken dialogs. In Nicoletta Calzolari, et al., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Wan, S., Kutschbach, T., Lüdeling, A., and Stede, M. (2019). RST-tace a tool for automatic comparison and evaluation of RST trees. In Amir Zeldes, et al., editors, *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019*, pages 88–96, Minneapolis, MN. Association for Computational Linguistics.

Zimmermann, M. (2011). Discourse Particles. In P. Portner, et al., editors, *Semantics*, number 33 in Handbücher zur Sprach- und Kommunikationswissenschaft HSK 2, pages 2011–2038. Mouton de Gruyter, Berlin.

Collecting Metadata for Social Media Corpora in the Face of Ever-changing Social Media Landscapes

Egon W. Stemle, Alexander König, Lionel Nicolas
Eurac Research, CLARIN ERIC, Eurac Research
egon.stemle@eurac.edu, alex@clarin.eu, lionel.nicolas@eurac.edu

Abstract

Social media is an essential part of people's lives, and communication has become increasingly participatory, interactive, and multimodal. The FAIR principles are essential for digital preservation and archiving, with Findability & Accessibility already well covered technologically. To ensure Reusability to future researchers and other stakeholders, documenting the platform itself is an essential part of metadata collection during corpus collection. A network of interested stakeholders should help develop standards and best practices for documenting, collecting, and archiving social media data.

Keywords: FAIR, metadata, CMC, social media,

1. Introduction

The "FAIR Guiding Principles for Scientific Data Stewardship" (Wilkinson et al., 2016) have become a cornerstone of data management, promoting Open Science by making resources Findable, Accessible, Interoperable, and Reusable. In particular, *F*indability and *A*ccessibility can be easily achieved by storing data in a certified repository (Frey et al., 2020). However, to realize the full potential of *I*nteroperability and *R*eusability, domain-specific knowledge and domain relevance are key. In our domain, we believe the "CMC-core: a schema for the representation of CMC corpora in TEI" (Beißwenger & Lüngen, 2020) has made big strides forward for *I*nteroperability. However, *R*eusability could benefit from inspiration from Archival Studies and others.

The definition of archives has evolved over time. The traditional definition reflects a passive, unbiased approach, while a modern definition challenges this idea and states that as a "result of the prevalence of information and communication technologies" (Li, 2023), there is now an understanding that “the archivist actively shapes the archive” and that “various contextualties [...] are relevant to archival work.” (Lane & Hill, 2010)

In relation to CMC and social media, this perspective is relevant to understanding the meaning and importance of the provenance of the data on which our research is based. We usually consider it enough to state that 'the data' was collected from this or that social media platform – and if we want to be particularly thorough, or the platform provides the data already, we also mention the time of collection. Admittedly, some platforms provide further metadata, such as location (more or less precise), device type (like phone, tablet or desktop) and others.

However, social media platforms are evolving and rapidly changing, and so is the technology embedding and surrounding them. This, in turn, directly impacts how people interact with and operate the technology and social media itself. We should consider this more when interpreting data from social media. Ultimately, however, this means we must first and foremost recognise and record this information.

In the following we will first describe the problem we see in the fluidity of the shape of social media platforms and the technology used to interact with them. Then we will present some suggestions on how to solve this problem by providing additional metadata specifically describing the social media environment. And finally we end with some conclusions and a call to the CMC community.

2. The Problem

Social media environments have a large variance, only more so the longer this form of media exists and evolves. The environment is important to understanding social interactions and the language production situation. Knowing that a certain communication item is “a tweet” or “a chat message” is not enough detail.

2.1 The Platforms

One of the primary features of social media data is its contemporary form, which reflects the current trends and conversations of society and sets it apart from most other language data. Additionally, social media data is incredibly dynamic since social networks are continually evolving to meet the needs of users and the market. It is also worth noting that some social media platforms may not be as prevalent as others. Furthermore, social media data is unique since it is incredibly diverse and each platform has its characteristics, making it challenging to track and analyze across the board. This is, among other factors, due to the fact that social networks are often subject to the whims of their owners, who may make changes that affect the metadata and data accessibility.

Some platforms were only widespread within a certain group, region, age group, etc. and were otherwise little known. For example,

- Hyves was focused on the Netherlands (available in Dutch and English) between 2004 and 2013. At its peak, 10.3 million accounts were registered - with the whole population of the Netherlands at the time amounting to roughly 16 million.
- Orkut was one of the earliest social networking platforms, launched in 2004 by Google and dissolved in 2014. It gained significant popularity in countries like Brazil and India.
- StudiVZ was a German social networking platform founded in 2005. Its name stands for "Studentenverzeichnis," which translates to "student directory" in English. It gained significant popularity in Germany, Austria and Switzerland, similar to how Facebook initially

targeted college students in the United States. In the end, it struggled to keep up with Facebook's growth and dominance in the social media landscape. It was effectively dead after 2017 and officially dissolved in 2022.

- VKontakte, often abbreviated as VK, is a social networking platform founded in 2006. It is commonly referred to as the "Russian Facebook" due to its popularity in Russian-speaking countries and similarities in functionality and layout. It remains a prominent social networking platform in Russia and neighboring countries.

Other once well-known, but now (de facto) defunct platforms include Myspace (which has been online since 2003), Google+ (2011 - 2019), or Geocities (1994 - 2009); and the next social media are already waiting in the wings for the near future, like Bluesky, Threads or Mastodon.

If we take a quick look at a currently established service (like X, formerly Twitter), there are a few things to mention that changed over the course of its existence. For example, the maximum length of a tweet changed over time, as did the method how this maximum length was determined (whether e.g. usernames or URLs counted toward this limit). Another example would be how URLs used in a tweet were handled: whether a preview was generated, whether the original source was clearly visible or whether the URL was automatically shortened. And this does not even include any mention of the turmoil around the change from Twitter to X - and all things related to this, e.g. changes in moderation and verification policy, or technical issues like outages.

Looking at another prominent social media site, the evolution of Facebook's posts and content types can be categorized into several key stages, starting from text-only status updates in 2004 to the emphasis on visual content and prioritizing posts from friends and family over public content from brands and publishers in recent years. As Facebook gained popularity, the platform introduced support for multimedia content such as photos and videos. Users could now enrich their posts by including images or sharing videos directly within their updates. Later, Facebook introduced the ability to "Like" and comment on posts, enhancing the interactive nature of posts. Additionally, users could share links to articles, websites, and other external content, expanding the types of content that could be shared on the platform. Facebook continued integrating new features into posts, including the ability to tag friends, check in at locations, and share feelings or activities using predefined emojis and status options. The introduction of multimedia content, expansion of content types, integration of new features into posts, and greater emphasis on visual content, such as photos, videos, and Stories, were among the significant changes in Facebook's evolution.

Finally we also have to mention the trend towards more video-focused platforms in recent years, be they TikTok, Instagram reels or Youtube shorts. Also within the CMC scientific community these platforms have gained more and more prominence recently and they should not be overlooked when documenting social media platforms.

2.2. The Technology

Significant advancements in processing power, memory capacity, and portability have marked the development of computer hardware from the 1990s to today. The impact of ultramobile devices like smartphones and tablets on social media platforms has been profound. These devices have facilitated anytime, anywhere access to social networking sites and apps, leading to increased user engagement and the proliferation of user-generated content.

In their early years, personal computers (PCs) were big and expensive, with limited processing capabilities and memory. However, over the years, there has been a consistent trend towards miniaturization and increased performance.

In the 1990s, PCs were the primary computing devices available. These devices were typically bulky desktop towers with CRT monitors. While laptops existed, they were expensive and not as common.

The early 2000s saw the rise of laptops as viable alternatives to desktop PCs. Improvements in battery technology and processors made laptops more powerful and portable. Additionally, advancements in graphics processing units (GPUs) led to better multimedia capabilities.

The mid to late 2000s saw the emergence of smartphones and tablets, which began to reshape the computing landscape. Smartphones like the iPhone and Android devices introduced touchscreens, mobile internet connectivity, and an ecosystem of mobile apps. Tablets, popularized by the iPad, offered larger touchscreens and enhanced portability.

Throughout the 2010s, mobile technology rapidly evolved. Smartphones became ubiquitous, with increasingly powerful processors, high-resolution displays, and advanced cameras. The rise of affordable mobile networks facilitated faster internet speeds, enhanced connectivity and widespread adoption. Tablets continued to improve, catering to both consumer and enterprise markets.

The evolution of interaction with digital devices, particularly in terms of input methods, has undergone significant transformations over the years as well.

In the past, users heavily relied on physical keyboards to input characters into their digital devices. This traditional method dominated the early days of computers. On mobile phones, "Text on 9 keys" (T9) was an early technological advancement to help users input text. It was a predictive text technology commonly used in mobile phones with numeric keypads. As an early form of modern autocorrect prediction technology it superseded the more cumbersome technique that required users to press keys multiple times to cycle through letters and form words.

However, the introduction of touch screens completely transformed how we interact with our digital devices. With touch screens, users can now directly input commands and text by tapping or swiping on the screen without needing physical keys. Predictive writing has become an essential feature of touchscreen keyboards on smartphones and tablets. It anticipates the word a user intends to type and suggests it as they enter text. This

technology leverages algorithms to predict words based on the context of the sentence, frequently used words and other factors.
Finally, today, integrating artificial intelligence (AI) and machine learning (ML) technologies into mobile devices has become prominent. Smartphones and tablets now feature AI-powered virtual assistants that include voice recognition, advanced camera functionalities, and personalized user experiences with independent task execution. All of these can suggest and generate much more than traditional predictive writing systems.

3. Suggestions: Going Meta on Metadata

We propose *Going Meta on Metadata*[1] by adding a section to the metadata or description of the corpus that provides information about the specific environment in which the data was collected. For example, rather than simply stating that the corpus contains Twitter data, corpus collectors should include more details allowing to understand what Twitter looked like at the time the data was collected and how it could have potentially influenced the production of the data (e.g. by limiting the size of inputs provided by users). The specific information to be included should be discussed within the community. The following paragraphs contain some first examples that could start such a discussion.
To capture user experience on a social media platform in greater detail, it is important to take note of the social graph, that is, how users can "follow" other users on the platform. What is the entry point for a user, and does a tailor-made "for you" page serve as an entry point? Can users only see messages from people they have pre-selected? Furthermore, does the platform show additional topics-related information in a side column? Lastly, are there global, local, and/or personal trends that users might refer to while using the platform?
What kind of interaction does the platform support? For example, something like retweets and quote tweets can each serve different purposes, where one is a 1:1 copy of the original, and the other contains the original content and some additional commentary. Another important aspect is whether threading is possible, common, or even encouraged on a particular platform. Threading involves creating a series of connected texts in response to one another. Finally, it is also important to consider whether there is moderation on a given platform and whether it follows clear and documented standards. All of these factors can impact how people interact on social media and the norms that emerge within a given community.
Which items such as text, photos, videos, or audio does the platform support? When it comes to entering text, does the platform include emojis, emoticons, gifs, or stickers as a way for users to enhance their content? Does the platform enforce character limits? Lastly, does the platform support screen readers or other means to accommodate users with accessibility needs?
We acknowledge that there are also additional "hidden features" that influence the user experience on social media, namely recommendation mechanisms and automatic or manual moderation, features which in the media are often subsumed as "the algorithm". Unfortunately, these features are usually completely undocumented and intransparent and we therefore do not think they should or could be part of our proposed metadata section.
Social media corpora encompass a vast range of services, each with its unique characteristics and features. While Twitter was once the go-to service for collectors of social-media corpora, it is no longer as easy to use, which means that the variety of social media corpora is expected to grow even further. Environments can differ significantly from one service to another, and they can also change over time, even within the same service. It is important to note that services can die out and be forgotten, which makes it crucial to include detailed documentation that covers the service in question.

4. Conclusion

In this article, we raised awareness on the concept of "Going Meta on Metadata" for documenting metadata focussing on the social media platforms used to generate the data included in a CMC corpus. However, since metadata standards are not imposed on a community but rather emerge and evolve through collaborative efforts within the community itself, we think a network of interested stakeholders should help develop standards and best practices for documenting, collecting, and archiving social media data. Ideally, once such a framework has been developed, it could then be used by default in all newly created CMC corpora and maybe even added as extended metadata to existing corpora, if possible.
Our objective is to foster a collaborative environment where the CMC community actively contributes to shaping the standards to meet the specific needs of the field. The CLARIN K(nowledge)-centre for Computer-Mediated Communication and Social Media Corpora (Stemle et al., 2022) could likely play the role of a coordinating entity.

5. References

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

Frey, J.-C., König, A., Stemle, E., Falaise, A., Fišer, D., & Lüngen, H. (2020). The FAIR Index of CMC Corpora. In J. Longhi & C. Marinica (Eds.), *CMC Corpora through the prism of digital humanities*. L'Harmattan. https://api.zotero.org/users/332053/publications/items/R7Z5VHA2/file/view

Kramer, M. J. (2014). Going Meta on Metadata. *Journal of Digital Humanities*.

[1] This phrase is inspired by "Going Meta on Metadata" (Kramer, 2014), where the phrase means something different – but the concept applies.

https://journalofdigitalhumanities.org/3-2/going-meta-on-metadata/

Lane, V., & Hill, J. (2010). Where do we come from? What are we? Where are we going? Situating the archive and archivists. In J. Hill (Ed.), *The Future of Archives and Recordkeeping* (1st ed., pp. 7–26). Facet. https://doi.org/10.29085/9781856048675.002

Li, M. (2023). From Archive to Anarchive: How BeReal Challenges Traditional Archival Concepts and Transforms Social Media Archival Practices. *Journal of Contemporary Archival Studies*, *10*(1). https://elischolar.library.yale.edu/jcas/vol10/iss1/9

Stemle, E. W., Frey, J.-C., König, A., Falaise, A., Erjavec, T., & Lüngen, H. (2022). Introducing the CLARIN K(nowledge)-Centre for CMC and Social Media Corpora (CKCMC). *Book of Abstracts of the 9th Conference on Computer-Mediated Communication (CMC) and Social Media Corpora (CMC2022)*, 46–47. https://hdl.handle.net/10863/37043

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018–160018. https://doi.org/10.1038/sdata.2016.18

Talking to oneself in CMC: a study of self replies in Wikipedia talk pages

Ludovic Tanguy[1], Céline Poudat[2], Lydia-Mai Ho-Dac[1]

1: CLLE - CNRS & University of Toulouse, France
2: BCL - CNRS & University of Nice Côte d'Azur, France

Email: ludovic.tanguy@univ-tlse2.fr, celine.poudat@univ-cotedazur.fr, lydia-mai.do-dac@univ-tlse2.fr

Abstract

This study proposes a qualitative analysis of self replies in Wikipedia talk pages, more precisely when the first two messages of a discussion are written by the same user. This specific pattern occurs in more than 10% of threads with two messages or more and can be explained by a number of reasons. After a first examination of the lexical specificities of second messages, we propose a seven categories typology and use it to annotate two reference samples (English and French) of 100 threads each. Finally, we analyse and compare the performance of human annotators (who reach a reasonable global efficiency) and instruction-tuned LLMs (which encounter important difficulties with several categories).

Keywords: Wikipedia talk pages, self reply, monologues, annotation

1. Introduction

Wikipedia Talk pages have been extensively studied as they provide a unique means to examine the dynamics of interaction between users in their collaborative efforts to contribute to the online encyclopaedia (Laniado et al. 2011, Gomez et al. 2011, Lüngen & Herzberg 2019, Schneider et al. 2010, Kopf 2022).

In this study, we propose to focus on a very specific and understudied practice in Wikipedia talks and written CMC studies in general: monologues in interactional settings i.e. situations in which users persist in posting even when no one else is intervening. Such situations are close to the phenomenon of "no response", which would not be uncommon in other written CMC genres such as forums for instance - Beaudouin and Velkovska were already observing in 1999 that 41% of the posts in a forum had not received any response within one month. Monologues in interactional settings are quite different from other situations which have been described so far, such as teacher monologues in the classroom or in vlogs for instance (e.g. Frobenius 2011). Indeed, these situations are intrinsically monological, which is not the case in Wikipedia talks.

To explore this phenomenon, we decided to start from the beginning, and concentrate on monologue inceptions. To achieve this, we analyse thread onsets in which users technically reply to themselves in a thread they have initiated. We focus on the reasons why a user would post a second message to himself/herself. We first provide an overview of this phenomenon in Wikipedia and examine the specific lexicon found in the posts. We then propose a typology of the seven main reasons we have identified for a self reply. We asked human coders to apply this typology on a sample of 100 threads. We analyse the results and finally present a first try at using Large Language Models to annotate the large amount of data available and discuss the overall difficulty of the annotation task.

2. Corpus and phenomenon overview

We base our study on the *EFGCorpus*, a comparable corpus composed of all the talk pages extracted from the August 2019 dump of the English, French and German versions of Wikipedia (Ho-Dac, to appear).

All the talk pages included in the EFGCorpus are encoded according to the TEI CMC-core schema (Beißwenger and Lüngen 2020). For each language, we selected all the threads in which there are neither unsigned and undated messages nor messages signed by a bot.

As Wikipedia talk pages are written the same device as the article pages, and therefore allow for a loose structure, we considered the linear order of messages as they appear in each thread. Table 1 gives an overview of the corpus and details concerning consecutive posts by the same author. All three languages show similar tendencies, although the quantities vary according to the size of the Wikipedia communities.

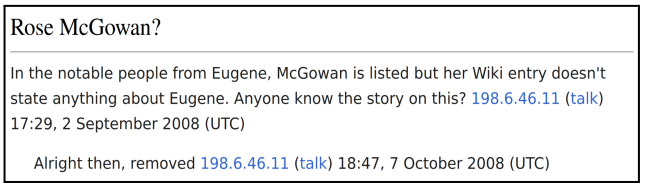
Feature	English	French	German
Number of threads	3 385 583	302 475	1 485 648
Number of posts	8 873 620	769 880	3 967 726
Threads with 2 posts or more	1 688 939	140 904	784 605
Threads containing two consecutive posts by the same author	406 292 (24,1%)	38 706 (27,5%)	179 871 (22,9%)
Threads starting with two consecutive posts by the same author	201 280 (11,9%)	19 947 (14,2%)	82 629 (10,5%)
Single-author threads with 2 posts or more	115 813 (6,9%)	12 019 (8,5 %)	48 413 (6,2%)

Table 1: Quantitative data overview

Half of the threads in the corpus contain only one post and thus are not conversations, whether because they have failed (no one answered the user who left the discussion as it is) or do not require a response (e.g. the post contains simple information). Among the subset of threads which contain at least two messages, about 20-25% contain at least one pair of consecutive posts by the same author. We decided to focus on an even more specific subset by selecting the threads which begin with two messages by the same author. This phenomenon is clearly very frequent as these represent more than 10% of the threads with two or more messages. Finally, we can see that 6-8% of all threads in our corpus are purely monological, which is significant.

Figure 1 shows an example thread in English in which the author of the first message (*198.6.46.11*) replies in the second message[1].

[1] Unregistered Wikipedia users are identified by their IP address. Therefore it is possible that we missed a number of self replies, as a user's IP can change between two posts. On the other hand, false positives are very highly improbable.


Rose McGowan?

In the notable people from Eugene, McGowan is listed but her Wiki entry doesn't state anything about Eugene. Anyone know the story on this? 198.6.46.11 (talk) 17:29, 2 September 2008 (UTC)

Alright then, removed 198.6.46.11 (talk) 18:47, 7 October 2008 (UTC)

Figure 1: Example single author thread
https://en.wikipedia.org/wiki/Talk:Eugene,_Oregon/Archive_2#Rose_McGowan?

In this example, the user opened a thread in order to highlight a lack of information or coherence and replied one month later to his first message using an interactional marker (*Alright then*) in order to report that he solved the problem by deleting the aforementioned section in the article. We estimated that focusing on self reply at the beginning of a thread is a more direct approach to the phenomenon, and also easier to investigate, as it does not require to follow a sometimes lengthy discussion.

We will now concentrate on our major research question: what are the reasons why the user posts a second message as a reply to his/her first one?

3. Self replies in Wikipedia - particularities and motivations

Monologues have been mostly studied in speech contexts, like in classroom interactions (see for instance Davis 2007) or video blogs or vlogs (Frobenius 2011, 2014) in CMC situations. Nevertheless, they have not been studied in written CMC interactions to our knowledge.
We propose to characterise monologue onsets in Wikipedia talks. After a quick glance at the most significant lexical specificities of the second message compared to the first, we report an annotation experiment aiming at categorising the main reasons why the user has posted a second message in a thread he initiated.

3.1 Keyness analysis of second messages

As shown in Table 2, the user's second message is significantly distinct from the first they posted, while it explicitly refers to it. Users are completely aware they are posting a second message, as we can observe with the words *PS*, *p.s.*, *update* or *forgot*, which explicitly complement or rectify their initial post. The use of the present perfect, which has an evaluative value, is also noteworthy (***I've** done / fixed / found / removed*…), as the Wikipedians report something they have done following their initial message - something they had possibly requested to be done, but which was still undone.

Word	Keyness	Word	Keyness
've	1000	ah	244
above	1000	mind	227
ahead	1000	also	226
done	1000	added	221
fixed	1000	you	215
I	1000	update	207
nevermind	1000	response	187
now	1000	oops	169
oh	1000	went	147
OK	1000	!	140
okay	1000	update	136
PS	308	further	134
still	293	nobody	133
again	284	objections	133
p.s	282	may	131
sorry	269	forgot	119
found	265	no	119
removed	262	thanks	119

Table 2: Main specificities[2] of the user's second message

We also note that classical interaction marks are present - although users are technically replying to themselves, the speech remains explicitly addressed, with the use of *you*, *nobody* or even *objections*. Even more strikingly, the user appears to utilise interactional markers like *OK* for instance to respond to themselves. Even in a monologue, the framework of the conversation seems to prevail.

3.2 A typology of self replies

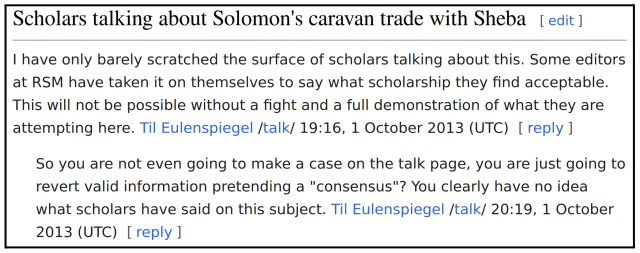
We carried out a detailed examination of the content of the threads beginning with a self reply, randomly selected in the English and the French parts of the xxxCorpus. After a first exploratory stage during which two of us annotated 200 threads (100 in English, 100 in French), we established a first typology of the reasons why users reply to themselves. This typology was then finalised in a second step of curation. As a result, 7 main reasons have been identified and defined as follows:

- **Addendum:** the user complements their first message with new information, a new scope, additional arguments or some kind of clarification;
- **Self-correction**: the user has identified an error in their first message and corrects it, possibly cancelling the first message;
- **Self-answer:** the user answers the question they asked in the first message;
- **Chasing up**: having received no replies to his first message, the user asks other users for answers or reactions;
- **Action report**: the user has done something since their first message and announces it;
- **Reaction to event**: something has been done by someone else, or has happened since the first message and the user reacts to this event;
- **List**: the first two messages constitute a list of items or the beginning of a list; these items can be pieces of information, things to do, remarks, questions etc.

If the first four categories were expected and could be observed in other online discussion platforms, the last three seem more specific to Wikipedia talk pages. **Action report** is crucial in the context of collaborative working. Suggesting, requesting or validating an article edit are amongst the main reasons why users chat in talk pages, as stated in Ferschke (2014) who showed that around 60% of the messages are associated with explicit performative speech acts. As shown previously, the user who requests an action and the user who performs it may be the same. In this situation, the content of the second message may be reduced to a minimum (*Done*) or it may contain details of the action performed (as in Figure 1).
The **Reaction to event** category differs from the others in that it goes beyond the framework of mere discussion. In Wikipedia, interactions may indeed spread over multiple

[2] The index is based on a calculation grounded on the hypergeometric distribution, using the *textometry* R package (https://cran.r-project.org/package=textometry).

channels. In this category, the user writes a second message in reaction to what someone else has done in another space e.g. mostly an article edit or sometimes a message in another discussion channel. Such cases are often difficult to understand because of the lack of contextual information, as in Figure 2 where *Til Eulenspiegel* addresses his second message to someone who "*revert[ed] a valid information*", using a second-person address[3] - this may also explain the prevalence of *you* in the keyness analysis.

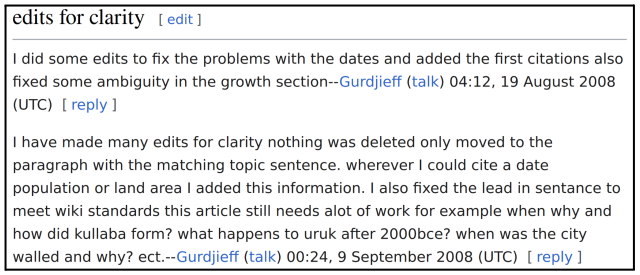

Scholars talking about Solomon's caravan trade with Sheba [edit]

I have only barely scratched the surface of scholars talking about this. Some editors at RSM have taken it on themselves to say what scholarship they find acceptable. This will not be possible without a fight and a full demonstration of what they are attempting here. Til Eulenspiegel /talk/ 19:16, 1 October 2013 (UTC) [reply]

So you are not even going to make a case on the talk page, you are just going to revert valid information pretending a "consensus"? You clearly have no idea what scholars have said on this subject. Til Eulenspiegel /talk/ 20:19, 1 October 2013 (UTC) [reply]

Figure 2: Example of a **Reaction to event**
https://en.wikipedia.org/wiki/Talk:Sheba#Scholars_talking_about_Solomon's_caravan_trade_with_Sheba

The **List** category is usually found in long monologues in which the users uses a thread as a logbook, a dashboard or a personal to-do list[4]. In such threads, the user just wants to keep trace of a work in progress without any intention of calling on the intervention of another user (cf. the self-commitment category in Ferschke 2014). In Figure 3, *Gurdjieff* lists all the edits he did on the 'Uruk' article.


edits for clarity [edit]

I did some edits to fix the problems with the dates and added the first citations also fixed some ambiguity in the growth section--Gurdjieff (talk) 04:12, 19 August 2008 (UTC) [reply]

I have made many edits for clarity nothing was deleted only moved to the paragraph with the matching topic sentence. wherever I could cite a date population or land area I added this information. I also fixed the lead in sentance to meet wiki standards this article still needs alot of work for example when why and how did kullaba form? what happens to uruk after 2000bce? when was the city walled and why? ect.--Gurdjieff (talk) 00:24, 9 September 2008 (UTC) [reply]

Figure 3: Example of a **List**
https://en.wikipedia.org/wiki/Talk:Uruk#edits_for_clarity

A last category had to be added: **Error** is used when there is only one message, when the first two messages are not written by the same author or are unrelated (i.e. do not belong to the same thread), which can be due to various factors (syntactic anomalies, noncompliance with editing conventions…).

3.3 Adjudicated dataset for English and French

Once this first typology was created on the basis of a first exploratory annotation by two of us, an adjudication phase allows us to provide a consensual dataset. Table 3 shows that in both languages, the two main reasons why a user writes a second message in a thread he has opened are to complement his first message (**Addendum**) or to report an action he did (**Action report**). Half of the annotated messages could be explained by these two reasons.

The third more frequent label is the **Error** category which means that 16% for English and 11% for French of our thread do not actually begin with a self reply (due to processing errors or specific configurations).
Self-correction, **self-answer** or **Chasing up** are the least frequent categories.

Categories	English	French
Addendum	30	24
Action report	24	26
Reaction to event	8	14
List	8	10
Self-correction	4	11
Self-answer	6	2
Chasing up	4	2
processing error	16	11
Total	100	100

Table 3: Adjudicated annotation of the second message for English and French

3.4 Human annotation and inter-annotator agreement

We asked two students in linguistics to apply the typology to the French and English dataset and measured the inter-annotator agreements between the adjudicated and the student annotations. The students spent around 3 hours each to annotate the 100 posts with a simple task: for each thread, they were asked to focus on each first and second message independently of the rest of the talk (i.e. whether there is a third message, by whom and for what reason), attempting to identify the second message's main function. We compared their annotations with the adjudicated categories described above and obtained Cohen's Kappa scores of 0.67 for French and 0.69 for English. We considered that this validates our typology and annotation guidelines.
Table 4 gives detailed F1 scores per category. The category with the highest agreement is **Action report** (F1=0.88 FR / 0.84 EN). For the French set, the three most divergent categories are the less frequent one i.e. **Chasing up** (F1=0.33), **Self-answer** (F1=0.50) and **Self-correction** (F1=0.59). It is also the case in English for the categories **Self-correction** (F1=0.54) and **Reaction to event** (F1=0.55). It has already been clearly demonstrated that the rarer a category is, the more difficult the item is to annotate (cf. Paun et al. 2022).

Category	Anotator 1 (French, k=0.67)	Annotator 2 (English, k = 0.69)	Mistral openorca (English, k=0.17)
Addendum	0.65	0.71	0.55
Self-correction	0.59	0.54	0.57
Self-answer	0.50	0.67	0.00
Chasing up	0.33	0.80	0.15
Action report	0.88	0.84	0.39
Reaction to event	0.67	0.55	0.17
List	0.60	0.57	0.00
Macro-average F1	0.60	0.67	0.28

Table 4: Detailed F1 scores per category for two human annotators and the best LLM.

In any case, the human annotators obtained much better scores than the LLMs, which we document in the next section.

[3]Note that this user was generically addressed as third-person (*some editors at RSM*, *they*) in the first message.
[4] Wikipedians are supposed to use dedicated "to-do" talk pages rather than the main talk pages for listing the things to do in the article, see https://en.wikipedia.org/wiki/Wikipedia:To-do_list.

3.5 Classification by Large Language Models (LLM)

It is a well-known fact that recent advances in NLP technology allow for efficient and flexible systems that can annotate text data for complex phenomena (Alizadeh et al. 2023). We wanted to estimate the difficulty of automatically classifying the monological thread beginnings, which would allow us to have a larger dataset and investigate the phenomenon further.

For this experiment we selected seven generic instruction-trained Large Language Models, limiting our choice to the smaller open-source models that could be run locally on a workstation with a small GPU (up to 13 billion parameters with quantization)[5]. We prompted these LLMs with a zero-shot approach (i.e. without examples) with the following instructions:

You are an expert linguist specialised in the study of online interactions. You will annotate online discussions from the Wikipedia talk pages where the same user replies to himself, and identify the main reason for this, using the following seven categories:

[Description of categories as in the bulleted list in § 3.2]

Below are the first two messages of a discussion (indicated by <MSG1> and <MSG2>). You will answer with the chosen category number for the second message, and only this number, without details nor explanation. You can decide that there is not enough data for answering and give a "NULL" answer.

Each thread in the English adjudicated dataset was processed independently and the answers had to be manually interpreted in most cases as they rarely respected the requested format. We compared these with the adjudicated annotations: we obtained Cohen's kappa scores ranging from 0 to a low maximum of 0.165 for Mistral-openorca (Lian et al. 2023). These scores clearly indicate a low efficiency of LLMs for this task, far below what our two students could achieve. The rightmost column in Table 4 gives the F1 scores for each category for the aforementioned best LLM. This particular system performed best on **Addendum** and **Self-correction**, approaching the students' scores. **Addendum** can be seen as a generic default answer, while **Self-correction** benefits from obvious linguistic cues as discussed above. On the other end, **Self-answer** and **List** categories could not be properly identified. However, this last type is clearly out of reach for the technique we considered, as a list cannot generally be detected on the sole basis of the first two messages. Human annotators had a clear advantage as they could access the whole thread and have a global view of the recurring pattern of messages. We cannot conclude for other categories when looking at the other LLMs we experimented, as each showed a very different behaviour.

4. Conclusion

We have proposed a first investigation of monological thread onsets in Wikipedia talk pages. This phenomenon is clearly quite frequent and interesting, and we have proposed a first typology of the seven main reasons why a user may reply to oneself. We obtained a satisfactory first trial with human annotation. Although we have no doubts that the use of LLMs would be extremely useful to annotate on a larger scale, this still needs further investigation and experimentation, notably to stabilise the categories.

We have a number of perspectives to investigate. First, we need to extend our human annotation, and we are currently enlarging our dataset. While this extension will enable us to establish a gold standard corpus, it will also allow us to identify specific cues that could be used as a pre-annotation. We will then be able to perform a first analysis of the feature for each category (length, global pattern, time profile, favoured topics or dialogue acts etc.).

On the other hand, this step is crucial to characterise longer monologues, extending the annotation to third, fourth messages and more. One of our ultimate goals would be to characterise types of monologues (entire threads) and monologue sequences (parts of threads), initially within Wikipedia, and eventually within other CMC genres.

[5] Mistral-Openorca 7b, Mixtral 0.2 7b, Gemma 7b, Mistral 8x7b, Llama2 7b and Llama2 13b. All models were run locally through the Ollama platform (ollama.com).

References

Alizadeh, M., Kubli, M., Samei, Z., Dehghani, S., Bermeo, J. D., Korobeynikova, M., & Gilardi, F. (2023). Open-source large language models outperform crowd workers and approach ChatGPT in text-annotation tasks. *arXiv preprint arXiv:2307.02179*.

Beaudouin, V., & Velkovska, J. (1999). Constitution d'un espace de communication sur Internet (forums, pages personnelles, courrier électronique...). *Réseaux. Communication-Technologie-Société*, 17(97), 121-177.

Beißwenger, M. & Lüngen, H. (2020). CMC-core: a schema for the representation of CMC corpora in TEI. *Corpus*, 20.

Davis, J. (2007). Dialogue, monologue and soliloquy in the large lecture class. *International Journal of Teaching and Learning in Higher Education*, 19(2).

Ferschke, O. (2014). *The Quality of Content in Open Online Collaboration Platforms: Approaches to NLP-supported Information Quality Management in Wikipedia*. PhD thesis, Technische Universität Darmstadt.

Frobenius, M. (2011). « Beginning a monologue: The opening sequence of video blogs ». *Journal of Pragmatics* 43: 814-27.

Frobenius, M. (2014). Audience design in monologues : How vloggers involve their viewers. *Journal of Pragmatics*, *72*, 59-72.

Gómez, V., Kappen, H. J., & Kaltenbrunner, A. (2011). Modeling the structure and evolution of discussion cascades. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia* (pp. 181-190).

Ho-Dac, L.-M. (to appear). Building a comparable corpus of online discussions in Wikipedia: the EFG WikiCorpus. In C. Poudat, H. Lüngen & L. Herzberg (Eds.), *Investigating Wikipedia: linguistic corpus building, exploration and analyses*. John Benjamins.

Kopf, S. (2022). *A Discursive Perspective on Wikipedia: More than an Encyclopaedia?* Palgrave Macmillan.

Laniado, D., Tasso, R., Volkovich, Y., & Kaltenbrunner, A. (2011). When the Wikipedians talk: Network and tree structure of Wikipedia discussion pages. In *Fifth international AAAI conference on weblogs and social media*.

Lian, W. Goodson, B., Wang, G., Pentland, E., Cook, A., Vong, C. and Teknium (2023). MistralOrca: Mistral-7B

Model Instruct-tuned on Filtered OpenOrcaV1 GPT-4 Dataset. *HuggingFace repository*. https://huggingface.co/Open-Orca/Mistral-7B-OpenOrca

Lüngen, H. & Herzberg, L. (2019): Types and annotation of reply relations in computer-mediated communication. *European Journal of Applied Linguistics* 7 (2). Berlin/Boston: de Gruyter, 2019. S. 305-331.

Paun, S., Artstein, R. and Poesio, M. (2022). *Statistical Methods for Annotation Analysis*. Morgan & Claypool publishers.

Schneider, J., Passant, A., & Breslin, J. G. (2010). A content analysis: How Wikipedia talk pages are used. In *Proceedings of the 2nd International Conference of Web Science* (pp. 1-7).

Who cares about correct spelling? Spelling discourse in social media conversations

Reinhild Vandekerckhove
University of Antwerp, Research group CLiPS
E-mail: reinhild.vandekerckhove@uantwerpen.be

Abstract

The present paper examines the discourse on spelling in a corpus of private social media conversations among Flemish teenagers on WhatsApp and Facebook Messenger. While explicit discussions on spelling issues are quite rare, the data reveal distinct gender differences: the adolescent girls express more attention for spelling issues than their male peers. However, a closer qualitative analysis of the data shows their commitment should not be overestimated: the adolescent girls tend to be more concerned with face-saving strategies than with a genuine commitment to correct spelling. From this ambiguity, it appears there is some (covert) 'coolness' linked to not prioritizing correct spelling. Finally, in line with previous research, we see that pointing out misspellings of others is deemed a face-threatening act that must be avoided. Clearly, there are no social benefits to be gained from acting as a 'spelling inspector' in private online interaction between peers.

Keywords: adolescent social media writing, spelling, face work

1. Introduction[1]

This paper builds on several studies of spelling practices in private social media conversations produced by Flemish adolescents who basically communicate in (Flemish) Dutch. Previous research did not only reveal that Flemish youngsters tend to produce a lot of spelling errors on homophonous verb forms that constitute notorious pitfalls in Dutch but also that some social groups are more prone to make mistakes than others: boys, younger teenagers and teenagers in technical/practical educational tracks produce significantly more errors than girls, older teenagers and teenagers in theoretical tracks (Surkyn et al 2020). Importantly, the misspellings of homophonous verb forms are 'real' errors that can be considered unintentional. They do not fit into the category of playful or functional deviations from standard spelling that are common in social media writing, since they are highly stigmatized among Dutch language users, especially because the spelling rules are clear-cut. Therefore, no prestige can be gained from producing them. Follow-up research in which adolescents' actual rule knowledge for the aforementioned verb forms was tested and adolescents were questioned about linguistic attitudes revealed that the gender difference could not be attributed to a difference in spelling proficiency, but had to be interpreted in attitudinal terms, with boys caring less about correct spelling (in social media contexts and beyond) than girls (Surkyn et al. 2022). Finally, an investigation into spelling correction practices by Flemish adolescents in private social media writing led to the conclusion that Flemish teenagers, regardless of their social profile, hardly correct their own spelling errors (of whatever kind[2]), or those of their peers (Surkyn et al. 2023). Generally, it appears hard to deduce whether they do not notice them or whether they do not want to perform a face threatening act or simply do not care. However, what remained out of scope until now is to what extent and how the adolescents in our corpus discuss spelling issues in their social media conversations. The present study wants to fill this gap by analysing the meta-level awareness and involvement of the adolescents as expressed in their social media discourse on spelling/writing both in a quantitative and qualitative way: to what extent do the conversations of the youngsters testify to a concern about spelling issues, do we see differences between social groups and is there any face work involved? The ultimate objective is to find out whether the explicit discourse on spelling corroborates our previous research, particularly concerning the observed gender patterns, and to identify what it adds in terms of the social and to some extent even psychological dynamics of norm compliance or the absence thereof in adolescent online peer group communication.

2. Data

The data are extracted from an anonymized corpus of private social media conversations produced on WhatsApp and Facebook Messenger in 2015-2016 by Flemish teenagers aged 13-20. The data were collected via secondary schools. On a voluntary basis, students donated conversations produced outside the school context and provided relevant metadata on their social profile (see table 1). The corpus consists of 456 751 posts (2 653 924 tokens). For more information about the data collection, we refer to Hilte et al. (2020a)[3].

[1] With sincere thanks to the anonymous reviewers for their highly relevant suggestions.

[2] This study no longer had an exclusive focus on the homophonous verb forms mentioned above. It included all spelling errors.

[3] Ethical clearance for data collection and secure data storage within research group CLiPS was given by the Ethical Advisory Committee for Social and Human

variable	levels	posts
gender	girls	301 189 (65.94%)
	boys	155 562 (34.06%)
age	13-16	245 709 (53.79%)
	17-20	211 042 (46.21%)
educational track	vocational	131 585 (28.81%)
	technical	204 617 (44.8%)
	theoretical	120 549 (26.39%)

Table 1: distributions in the entire corpus

From this corpus we extracted all posts that contained tokens of the Dutch verb *schrijven* 'write' (infinitive, inflected forms, past participle, and compounds with the verb stem *schrijf*, e.g. *schrijffout* 'writing mistake', *schrijfvaardigheid* 'writing proficiency'), of the verb *spellen* 'spell' (idem, including the compound *spelfout* 'spelling mistake'), of the noun *spelling* 'spelling' (and compound *spellingsfout* 'spelling mistake') and finally of the noun *typfout* 'typo' (the variant *tikfout* 'typo' yielded no hits)[4]. These searches rendered 1077 posts. In order to have some conversational context we added the two preceding posts and the post following the selected posts. Next a manual check was performed to select the posts which show personal concern or some sort of (positive or negative) attention for spelling issues. Only 133 posts (or 12.35% or the initial selection) met this criterion, because the verb *schrijven* 'write' was mainly used in contexts that were not spelling-related (e.g.: 'I am writing my text for tomorrow'). Importantly, the posts preceding and following the relevant posts are not included in this count. Generally, the selected posts address spelling issues that pop up during the online interaction, but in some cases the participants discuss spelling issues they are confronted with outside the online interaction (e.g. in a school task). As will be demonstrated below, these 133 posts (and their conversational context) reveal some interesting patterns, both from a quantitative and a qualitative perspective.

3. Results

3.1 Social patterns

First, we have to conclude that the adolescents hardly discuss spelling issues at all: in quantitative terms, the 133 relevant posts from 103 different participants constitute a nearly negligible portion of the total number of posts in the corpus, which obviously is most telling in itself: although the above-mentioned search does not cover all of the posts in which spelling issues are addressed (e.g. posts including straightforward corrections without any of the above-mentioned search terms are not included, see Surkyn et al. 2023), we can safely conclude spelling is no major concern. Still, some interesting patterns emerge, especially in terms of gender differences. 34 (25.56%) of the posts were produced by boys, 99 (74.44%) by girls. Clearly the skew in the representation between both gender groups is more pronounced than in the entire corpus (compare table 2 to table 1): if we relate the frequency of posts without reference to spelling issues to the frequency of posts with reference to spelling issues for both gender groups, the gender difference appears to be significant (X^2= 4.27, p = 0.039). Girls definitely express more concern about spelling than their male peers. Interestingly, we see no significant difference between younger teenagers and teenagers nearing adulthood: the representation of both groups in the spelling-posts is comparable to their representation in the entire corpus (X^2= 0.2, p = 0.65). Finally, students in theoretical educational tracks show more interest in spelling issues than their peers in the most practical (vocational) track (X^2= 9.42, p = 0.002), which may be related to a stronger focus on spelling in their school curriculum (see also Surkyn et al. 2022). The differences between the other groups (theoretical versus technical or technical versus vocational) however are not significant.

The attested gender difference aligns with previous findings regarding adolescents' actual spelling behavior in online contexts (Surkyn et al. 2020), and, by extension, regarding their broader linguistic norm compliance: adolescent girls were found to integrate fewer substandard speech markers in their social media writing than boys (Hilte et al. 2020b). This implies that boys' writing tends to deviate stronger from Standard Dutch not only in terms of spelling, but also in terms of grammar and lexicon. Therefore, it should not come as a surprise that, at a meta-level, girls express more concern with normative language issues such as correct spelling. The greater emphasis on spelling issues observed in the discourse of students from theoretical educational tracks also corresponds to their actual linguistic behavior: their online writing exhibits fewer spelling errors and substandard speech markers compared to that of their peers from the other tracks (Surkyn et al. 2020, Hilte et al. 2020b). However, this corroboration is not observed in terms of age: the informal social media writing of older teenagers deviates less from the standard norm than that of their younger peers (Surkyn et a. 2020, Hilte et al. 2020b), yet this is not reflected in a higher frequency of explicit discourse on correct writing.

variable	levels	spelling-posts
gender	girls	74.44%
	boys	25.56%
age	13-16	51.88%
	17-20	48.12%
educational track	vocational	18.80%
	technical	45.11%
	theoretical	36.09%

Table 2: distributions in the selection of 'spelling-posts' from the corpus

When focusing on the contents of the posts, gender again

Sciences of the University of Antwerp.

[4] With sincere thanks to Lisa Hilte for extracting the data

surfaces as a most interesting variable, while we do not see distinct patterns for the other social variables (i.e. age and education). Therefore, henceforth we will focus on this variable.

While some boys explicitly oppose to prescriptive spelling approaches, none of the girls does.

E.g.:

(1) *fuck it als da verkeerd geschreve is* 'fuck it if this is misspelled' (16 year old boy)

Some of the girls seem genuinely concerned with correct spelling, even when the tone remains light-hearted:

(2) *k zweer het gebruik gewoon correcte grammatica en spelling als je met mij stuurt hahahaha* 'I swear it, just use correct grammar and spelling if you interact with me hahahaha' (16 year old girl)

The following conversation between two girls aged 18 and 19 also combines a genuine concern (i.e.: using sms writing conventions is perceived to have a negative impact on spelling skills) with a touch of humor and playfulness. Girl 1 takes up the role of the wise advisor. While she seems to be serious about the content of her message, the end of the conversation shows both participants are aware of the playful character of this 'role-play'.

(3)

Girl 1: *En pas op met uw sms-taal eh* 'watch your sms-language'

Girl 1: *& klassiek met K remember* 'and classic with C, remember'

Girl 2: *Jaah schat ik weet heb* 'yeah darling I know is'

Girl 2: *het* 'it' (corrects final word in previous post)

Girl 1: *Probeer daarom languit te schrijven als je via sociale media bezig bent* 'Therefore try to write in full if you interact via social media'

Girl 1: *Dan oefening je ineens met alles echt waar* 'Then you exercise everything, believe me'

Girl 1: *Toen ik vroeger sms-taal gebruikte had ik altijd een buis voor schrijfvaardigheid* 'When I used sms-language before, I always failed for language proficiency'

Girl 1: *Sindsdien vele minder omdat ik nu amper of althans geen afkortingen meer gebruik die ik vroeger wel gebruikte* 'Now this happens a lot less, because I hardly use the abbreviations anymore that I used before'

Girl 2: *Dobrze schat zal ik zeker doen xx* 'Thanks darling I will do this xx' (Dobrze = Polish)

Girl 1: *Mooi zo :)* 'Good girl'

Girl 2: 😂😂😂😂😂😂

The most interesting gender difference however relates to face saving strategies.

3.2 Face work

Face work is a prominent concept in linguistic politeness theory as developed by Brown & Levinson (1987). Their work elaborated on the seminal work of Goffman (1967) who introduced the notion of 'face' as referring to the public image individuals want to establish or the way they want to be seen by the people they interact with. Brown & Levinson (1987) distinguish between positive and negative face wants: positive face wants relate to individuals' desire to be appreciated, negative face wants to their desire for autonomy (e.g., people do not like to be instructed or reprimanded). As Beißwenger & Pappert (2019: 232), who study face work in online interaction, put it: "social interactions are about saving or losing face", since speech acts may be face-threatening or face-saving, either for one's own face or for that of the other.

This frame appears highly relevant when analysing the spelling-posts, since participants are clearly involved in face work when addressing spelling issues. In most cases they use face saving strategies to preserve their own reputation by anticipating on potential failure. In other words, they try to avoid face loss, by expressing uncertainty about the correct spelling of a particular word (examples 4 and 5) or – less often – by apologizing for a spelling mistake (examples 6 and 7). Strikingly girls do this much more frequently (in 52,52% of their spelling-posts) than boys (23,53%) ($X^2 = 8.59$, $p = 0.003$). A sentence that pops up very often in girls' conversations in several (Flemish Dutch) variants is *of hoe je dat ook schrijft*, literally 'or how you write this', meaning: 'whatever the (correct) spelling may be'.

E.g.

(4) *Jaa da's beter dan die dretlogs achtig gedoe of hoe ge da ook schrijft* 😂 'Yes this is better than that dreadlock-like fuss or how you have to write this' (16 year old girl)

(5) *pfff ik ben bij de tattooeur of hoe ge da ook schrijft hahah* 'pfff I'm at the tattooist or how you have to write this hahah' (18 year old girl)

By making a reservation about the spelling they use and by adding expressions of laughter these girls protect themselves from a face-threatening act, e.g. being laughed at or corrected when producing an incorrect spelling for 'dreadlocks' or 'tattooeur' (the latter spelling is actually fine in Dutch). However, a side-note can be made about this recurring face-saving strategy: while the participants seem to be insecure about the correct spelling and do not want to lose face, their concerns are not such that they make an effort to check the correct spelling. In other words, their messages are somewhat ambiguous: they use a kind of disclaimer to make clear that they are well-aware of (potentially) not producing the standard spelling, but at the same time they demonstrate that they are cool enough not to spend time on checking the correct spelling (see also Meredith & Stokoe 2014, who found that Facebook chatters generally do not invest time in checking spelling). The boy in example (1) above even makes this explicit: in stating he does not care at all about a potential misspelling, he justifies not having checked the spelling and prevents a potential face-threatening act of his interlocutors. In fact, these adolescents seem to rely on a kind of implicit contract: all parties involved know that spelling mistakes are no big thing in adolescent peer group chat. Yet some of them seem to be anxious of making a bad impression and this

insecurity clearly manifests itself more strongly in girls than in boys. However, in view of the low number of explicit statements on spelling and the type of disclaimers presented here, for most of the chatters the implicit contract clearly suffices as such. In other words, most chatters do not feel the need to add these types of disclaimers. They generally do not feel the need to apologize for spelling mistakes or typos either. Still, we do see some counterexamples, once again especially in girls' chat.

E.g.

(6) *Srry vr de typfoute*😊😱 'sorry for the typos' (15 year old girl)

(7) *Sorry voor de schrijffouten*😊 'sorry for the writing errors' (19 year old girl)

Finally, while the disclaimers discussed above take up more than half of the selected spelling-posts produced by girls, pointing out the mistakes of the addressee(s), either by focusing on a particular mistake or in more general terms, is highly uncommon: only 5 of the 99 spelling-posts (5.05%) produced by girls contain such a face-threatening speech act (FTA) which is at odds with both the positive and negative face wants of the other, since it might signal disapproval and encourage the addressee to admit or rectify their mistake. These type of FTA's have a low frequency in boys' spelling posts too, although they have a larger share than in girls' posts (5/34 or 14.7%). Interestingly one of the boys responds by reproaching his interlocutor of being a spelling inspector, by addressing him as *meneer de spellingcontroleur* 'mister spelling inspector', which confirms the perceived face-threatening character of these types of interventions. It should be noted that comments on the spelling practices of third parties not involved in the conversation are excluded here, as these constitute no FTA for the participants in the conversation.

Strikingly, in 3 of the 5 cases girls add one or more expressive markers to mitigate the FTA (see example 8). One of the 5 boys 'remedial' posts also contains a FTA mitigator (example 9), in the other cases the mistake is pointed out in a straightforward way (example 10).

(8) *Kheb al een schrijffout gevonden int allerbegin*😊*xx* 'I found a writing mistake in the very beginning😊*xx*' (17 year old girl)

(9) *Wel nog een paar spellingsfoutjes haha*😝 'still some spelling mistakes haha😝' (16 year old boy)

(10) *da schrijfde wel met een d he* 'you have to write this with a d' (13 year old boy)

The use of emoji for mitigating FTA's is in accordance with the findings of Beißwenger & Pappert (2019) who studied online peer group feedback among students and found that critique and suggestions for improvement were often followed by emoji that thus served as politeness strategy devices. This is precisely what is happening in examples like 8 and 9 too.

In the end, the strategies discussed in the present section correspond to the essence of face work as described by Goffman (1967: 11) in his pioneering study: "The combined effect of the rule of self-respect and the rule of considerateness is that the person tends to conduct himself during an encounter so as to maintain both his own face and the face of the other participants".

4. Discussion

The analysis of the adolescents discourse on spelling issues confirms gender patterns from previous research: girls do not only produce less spelling errors in informal private online conversations (Surkyn et al. 2020) and they do not simply pay more lip service to correct spelling when explicitly questioned about it (Surkyn et al. 2022), they effectively express more attention and concern about spelling issues during their online interactions than their male peers. At the same time the analysis of the posts in which spelling issues are discussed reveals that their commitment should not be overestimated: the adolescent girls tend to be more concerned with face-saving strategies than with genuine dedication to correct spelling. Moreover, overall, these types of posts are scarce in the corpus, which obviously is most telling in itself. While girls tend to be care more, most chatters do not express any care about correct spelling at all.

Finally, in line with Surkyn et al. (2023), which focused on actual self- or other-corrections (so replacing the misspelled word with the correct equivalent), we see that pointing out someone else's mistakes is extremely rare. Adolescents surveyed by Surkyn et al. (2023) stated they considered correcting other's errors pedantic and annoying and therefore not done. While there is hardly any comparable research on this topic, a small scale qualitative study by Koh (2007) revealed a similar reluctance among younger children: she investigated nine children (aged 11) who participated in synchronous CMC chat sessions in a second language learning context. She concluded they either ignored their peers' mistakes or tried to point them out in an implicit way, because they did not want to embarrass or offend them. Apart from some exceptions, the present study shows that adolescents generally refrain from discussing the spelling behavior of their interlocutors in more general terms too. Suggesting to interlocutors that they should check their spelling clearly indeed is a face-threatening act that needs to be avoided or, when performed, has to be mitigated. No benefits can be gained from acting as a scrutinizer when hardly anyone cares about correct spelling and when not caring even seems to be deemed 'cool'. After all, you certainly do not want to be considered a *spellingscontroleur* (spelling inspector).

Understanding these socio-psychological dynamics can be an asset for educators and caretakers, helping to counterbalance the "moral panic about declining standards of literacy" (Thurlow 2006: 678) that often pervades the public discourse on youngsters' online writing practices. One could keep in mind that, to some extent, deliberately not paying attention to spelling and other writing issues in online peer group communication is part of adolescents' collective identity construction.

5. References

Beißwenger, M., Pappert, S. (2019). How to be polite with emojis: a pragmatic analysis of face work strategies in an online learning environment. *European Journal of Applied Linguistics* 7(2), special issue pp. 225—154

Brown, P., Levinson, S.C (1988). Politeness: Some universals in language use. Cambridge: Cambridge University Press.

Goffman, E. (1967). *Interaction Ritual: essays on face-to-face behavior.* New York: Garden City.

Hilte, L., Vandekerckhove, R. & Daelemans, W. (2020a). Linguistic accommodation in teenagers' social media writing: convergence patterns in mixed-gender conversations. *Journal of Quantitative Linguistics* 29(2), pp. 241—268.

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

Koh, Young-Ihn (2007). ESL children's error recognition and correction patterns in a synchronous CMC context. *English Teaching* 62(4), 257-278. Retrieved from: http://kate.bada.cc/wp-content/uploads/2015/02/kate_62_4_12.pdf

Meredith, J., Stokoe, E. (2014). Repair: Comparing Facebook 'chat' with spoken interaction. *Discourse & Communication* 8(2), pp. 181—207.

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. *Mental Lexicon* 15(3), pp. 422—463

Surkyn, H., Sandra, D. & Vandekerckhove, R. (2022). Adolescents and verb spelling : the role of gender and educational track in rule knowledge and linguistic attitudes. *Dutch Journal of Applied Linguistics* 11, pp. 1—19.

Surkyn, H., Sandra, D. & Vandekerckhove, R. (2023). When correct spelling hardly matters: teenagers' production and perception of spelling error corrections in Dutch social media writing. *European Journal of Applied Linguistics,* 2023, https://doi.org/10.1515/eujal-2022-0028

Thurlow, C. (2006). From statistical panic to moral panic: the metadiscursive construction and popular exaggeration of new media language in the print media. *Journal of Computer-mediated Communication* 11, pp. 667—701.

Language Style Accommodation in Computer-Mediated Communication: Alignment with Textisms, Emoji, and Emoticons

Lieke Verheijen
Radboud University (the Netherlands)
E-mail: lieke.verheijen@ru.nl

Abstract

This paper reports on a quantitative analysis of a corpus of experimentally elicited chat messages, to research language style accommodation (i.e., alignment, convergence) in informal computer-mediated communication. Native speakers of English and East-Asian speakers of English as a foreign language of different age groups (18-27, 40+) participated in the study, to explore any effects of native language and age on accommodation. Participants (N = 700) were presented with one-sided chat conversations that contained textisms (e.g., *wassup*, *thx*, *omg*), emoji (e.g., 😊👌☕), emoticons (e.g., ^^, B-), :-O), or none of the above. They were asked to respond to the chat messages; their responses were coded for the frequency of online language style elements to identify accommodation. The results indeed showed evidence of accommodation with textisms, emoji, and emoticons. The demographic variables of age and native language were not found to affect accommodation, but did affect participants' usage of textisms and emoticons.

Keywords: accommodation, alignment, convergence, textisms, emoji, emoticons

1. Introduction

1.1 Linguistic accommodation

Communication Accommodation Theory (CAT) postulates that in interacting with others in interpersonal and intergroup contexts, people often adapt their language style to match that of their conversation partner (Giles et al., 1991). Such accommodation (also called alignment or convergence) can reduce social differences and increase communication effectiveness, allowing interlocutors to understand each other better or to like each other more (Giles, 2016; Pickering & Garrod, 2006). In oral communication, speakers can align on different levels, both verbally (for example, through pronunciation, word choice, or syntax) and non-verbally (for example, with gestures).

1.2 Accommodation in CMC

Because of the ever-increasing popularity of social media, researchers have begun to explore the applicability of CAT to computer-mediated communication (CMC) or digitally mediated communication (DMC) (Brinberg & Ram, 2021, on texting; Danescu-Niculescu-Mizil et al., 2011, on Twitter; Muir et al., 2017, on instant messaging; Wang et al., 2009, on chatrooms). Some evidence of accommodation with linguistic style elements that are typical of written CMC ('digi-talk' or 'textese') has been found (Adams et al., 2018, 2023; Kroll et al., 2018; Marko, 2022; Siebenhaar, 2018; Wagner et al., 2022), but except for several studies on accommodation between Flemish youths by Hilte, Vandekerckhove, and Daelemans (Hilte et al., 2021, 2022, 2024), research in this field remains scarce. Moreover, much remains unknown about which factors affect accommodation in CMC.

1.3 Online language style

Written CMC has become highly multimodal. Accordingly, accommodation in CMC can take place with textual as well as visual elements. A textual element that is typical of written CMC are textisms, including non-standard abbreviations or 'phrase-shorteners' such as 'omg' (*oh my god*) and 'lol' (*laughing out loud*) and phonetic respellings such as 'cuz' (*because*) and 'tho' (*though*) (Thurlow & Poff, 2013; Adams & Miles, 2023). CMC is nowadays characterized by a plethora of visual elements, including pictures, GIFs, memes, stickers, emoji, and emoticons. The present study focuses on the latter two features, which can easily be inserted into written CMC on any platform. Emoji, which originated in Japan, are small ideograms that represent smileys, people, animals & nature, food & drinks, activities, travel & places, objects, symbols, and flags. Emoji can serve many functions, such as to compensate for a lack of non-verbal cues (😉👏🤷‍♀️), to add expressivity or emotion to text (😭❤️😁), to visualize or disambiguate messages (🎷🥪🧈), and to increase the informality or playfulness of online writing (💩🤡🤪) (Evans, 2017). Emoticons, the precursors of emoji, emerged in the United States and consist of typographic characters – letters, numbers, and punctuation marks. They share functions similar to those of emoji, as they can add sentiment to text (e.g., :-*, XD, ;-). One subtype of emoticons are kaomoji (e.g., ^_^), which should be read horizontally rather than vertically and which were popularised by East-Asian (specifically, Japanese) CMC users (Giannoulis & Wilde, 2019; Markman & Oshima, 2007).

2. Research aims

This study aims to find additional evidence of CAT in CMC and aims to explore if age group and native/first language (L1) affect accommodation. These variables were explored because (a) different generations may use CMC differently in part due to having grown up with CMC or not (cf. Prensky's (2001) 'digital natives' and 'digital immigrants') and (b) prior research by Wang et al. (2009) found an impact of culture or native language in communication style accommodation in CMC, with East-Asians aligning more than native speakers of English. Thus, a large-scale experiment was conducted to collect a chat corpus to investigate whether accommodation with textual and visual elements that typify written computer-mediated

communication takes place in the English written CMC of native speakers of English and East-Asian speakers of English as a foreign language (EFL) of different age groups.

3. Methodology

3.1 Design

A between-subjects experiment was designed with four conditions, including three online language style elements: textisms (e.g., *tbh*, *alrite*, *ttyul*), emoji (e.g., 😉🎲🤷), emoticons (e.g., =D, (^‿^), ;-p), or none. Three experimental groups were presented with chat scenarios that were developed to contain multiple instances of textisms, emoji, or emoticons, while a control group saw the same messages but without any textisms, emoji, or emoticons. Inspired by Kroll et al.'s (2018) methodology, participants were asked to type responses to the messages, which yielded a corpus that was quantitatively analysed for the relative frequency of textisms, emoji, and emoticons to identify accommodation, that is, convergence in participants' online language style to that of their imagined interlocutors. The research design thus includes the presence of online language style elements in the chat stimuli as independent variables, and the frequency of online language style elements in participants' responses, as well as participants'age and L1 as dependent variables.

3.2 Participants

700 people participated in this study. Regarding their native language, participants were either native speakers of English (n = 400) or East-Asian (Chinese, Japanese, Korean, or Vietnamese) EFL speakers (n = 300). As for their age group, participants were either classified as digital natives who had grown up with CMC (age range 18-27, Gen Z, n = 400) versus digital immigrants who learnt to use CMC at a later age (40 or older, mostly Gen X and Baby Boomers, n = 300). This study aimed for a balanced division of the four participant groups over the conditions, but since it proved difficult to recruit one of the participant groups, the following corpus was used for this paper:

Participant group	# participants	# chat responses
English younger	200	1,800
English older	200	1,800
East-Asian younger	200	1,800
East-Asian older	100	900
Total	**700**	**6,300**

Table 1: Corpus composition.

3.3 Materials

As experimental stimuli, three one-sided chat conversations were developed with at least four messages each. All messages contained questions, so as to elicit three responses per conversation from the participants. The tone of voice of the messages was informal and the topics (having drinks, meeting up for dinner, and someone's holiday) were intentionally casual, so that any textisms, emoji, or emoticons would not feel out of place. Four versions of each chat conversation were developed (so twelve stimuli in total): a control condition without non-standard orthography and visual features, as well as three experimental conditions with either textisms, emoji, or emoticons added. Other than the manipulation with these online language elements, the materials for the different conditions were identical. Each message included at least one online language element, but most included more (e.g., three emoji in one message), to make sure that the manipulation was salient. The conversations were visualized as WhatsApp chats and created with an online chat generator (Zeoob, 2024). The interlocutors were named 'Friend A.', 'Friend B.', and 'Friend C.' and their profile picture displayed a pet (dog, cat, and bunny respectively), so as not to indicate the interlocutor's gender, which may affect accommodation (Hilte et al., 2022). Examples of the materials are presented in the Appendix.

3.4 Data collection procedure

The corpus was collected using the online survey and recruitment platforms Qualtrics and Prolific. Only participants belonging to the L1s and age groups as specified in section 3.2 were selected. Participants were instructed to complete the study on their mobile device (i.e., smartphone), not on their desktop computer, so as to facilitate the addition of emoji to their responses, given the ease of using emoji keyboards on smartphones.[1] After completing the consent form, participants were asked a number of demographic questions. Next, they were exposed to one of the four conditions: they saw three chat conversations with textisms, emoji, emoticons, or none. Participants had to read the chat messages and write three responses for each of the conversations, resulting in nine responses per participant. They were explicitly requested to respond as they normally would to chat/text messages from actual friends. In addition, they were encouraged to not just answer 'yes' or 'no', but to give longer responses. At the end of the questionnaire, participants were asked about their ordinary frequency of use of textisms, emoji, and emoticons in their informal texts or chats. The average completion time was about 8 minutes.

3.5 Corpus coding

Participants' responses to the simulated chat messages resulted in a corpus of 6,300 responses. A codebook for manually coding the corpus for the frequency of textisms, emoji, and emoticons was developed; two coders were trained to use it. After a test round of coding and comparing coding differences between the two coders, the codebook was improved and finalized. One of the coders was designated main coder for the entire corpus; the second

[1] Note that desktop computers now also offer access to an emoji keyboard (Windows shortcut: logo key + period; Mac shortcut: Control + Command + Space), allowing emoji to be easily inserted by either browsing through an emoji keyboard or by typing in keywords. However, not all CMC users will be aware of these computer shortcuts.

coded a subset of the corpus (n = 300 participants; 2,700 responses) for computing intercoder reliability. Cronbach's alpha values for textisms, emoji, and emoticons were all excellent, with $\alpha \geq 0.95$. Two kinds of textisms were coded, namely word modifications and structural adjustments, the latter involving non-standard punctuation, capitalization, and diacritics. A second distinction was made between two kinds of emoticons, namely Western-style vertical emoticons and Eastern-style horizontal kaomoji. The corpus was coded using Microsoft Excel.

4. Results and discussion

In line with prior research, the presence of online language style elements was expected to be higher in the experimental groups' CMC than in the control group's CMC. This was confirmed, since significantly more textisms, emoji, and emoticons were used in the responses in the textism condition (F (3, 680) = 4.81, $p < .01$), emoji condition (F (3, 680) = 10.40, $p < .001$), and emoticon condition (F (3, 680) = 4.98, $p < .01$) than in the other conditions. These main effects are shown in Figures 1–3.

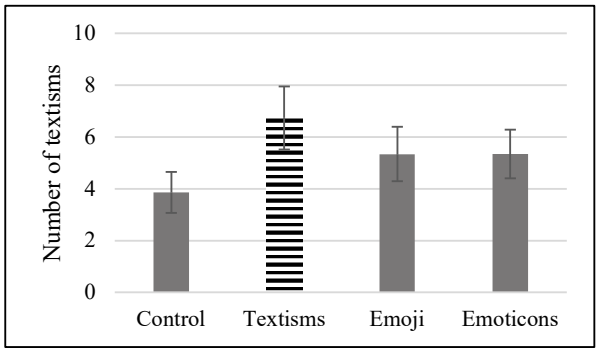


Figure 1: Textisms by Condition

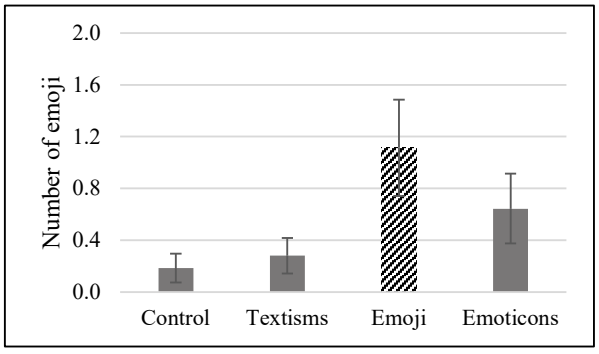


Figure 2: Emoji by Condition

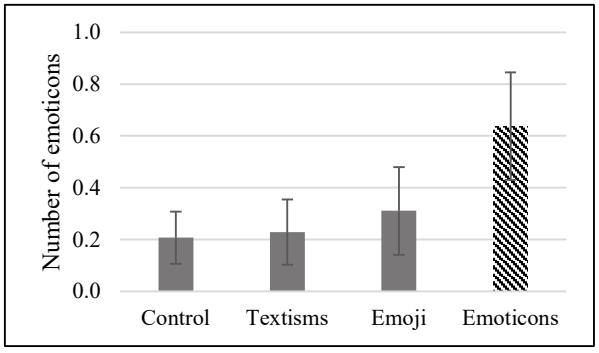


Figure 3: Emoticons by Condition

No evidence was found of age or native language affecting accommodation with online language style elements, as there were no interaction effects between Condition and Age group or Native language.

However, main effects and interaction effects revealed that age and native language did affect use of textisms and emoticons. As Figure 4 shows, textisms were used significantly more by younger participants than by older participants (F (1, 680) = 49.10, $p < .001$). This main effect may be attributed to the fact that the younger participants belonged to Gen Z, who – as digital natives – had grown up with CMC and texting abbreviations, while the older participants were digital immigrants who learnt to use CMC at a later age. Prior research has shown that younger people deviate more from standard language orthography, while older CMC users conform more to the standard spelling (e.g., Verheijen, 2017).

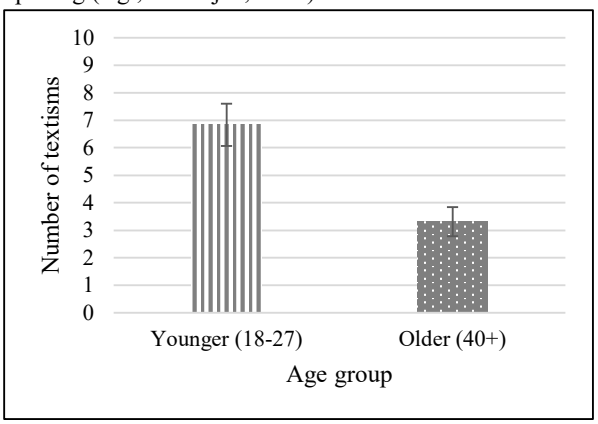


Figure 4: Textisms by Age group

Figure 5 shows that textisms were especially used more by younger East-Asian EFL participants (F (1, 680) = 5.67, $p < .05$). This interaction effect may be explained by IDLE, i.e., *Informal Digital Learning of English*: young people in Asia are among the digital natives who are increasingly acquiring the English language outside of the formal classroom by using digital tools (Lee & Dressman, 2018; Lee & Sylvén, 2021; Zhang & Liu, 2022). Such informal digital learning takes places through, for instance, social media interaction and online chat with native English speakers, in which textisms will be more prominent than in formal English education.

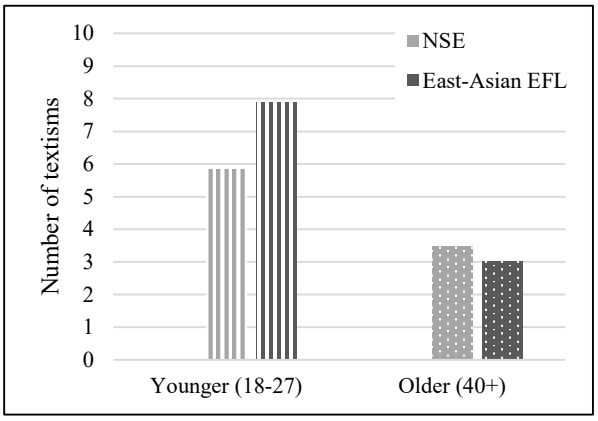


Figure 5: Textisms by Age group by Native language

The fact that no interaction effect between Condition and Age group was found indicates that there was no difference between Gen Z and older generations in the extent to which they align their language style to their interlocutors' use of textisms, emoji, or emoticons.

Figure 6 shows that emoticons were used significantly more by L1 English-speaking participants than by East-Asian EFL participants (F (1, 680) = 5.77, p < .05). A possible reason for this finding is that East-Asian people may prefer other visuals such as stickers, memes, or GIFs, over emoticons (Lu & Kroon, 2024).

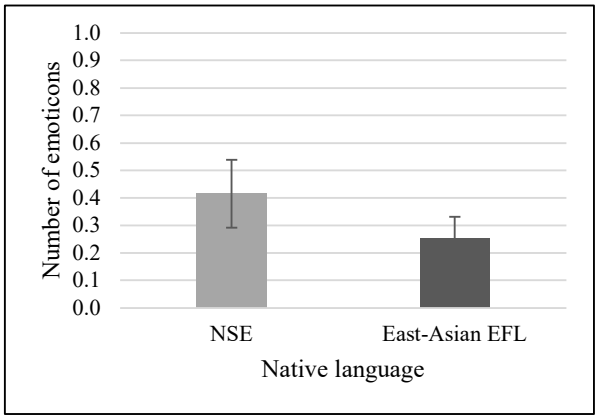


Figure 6: Emoticons by Native language

Older L1 English-speaking participants in particular were shown to prefer the 'old-fashioned' emoticons as compared to other participant groups (F (1, 680) = 4.44, p < .05), as visualized in Figure 7.

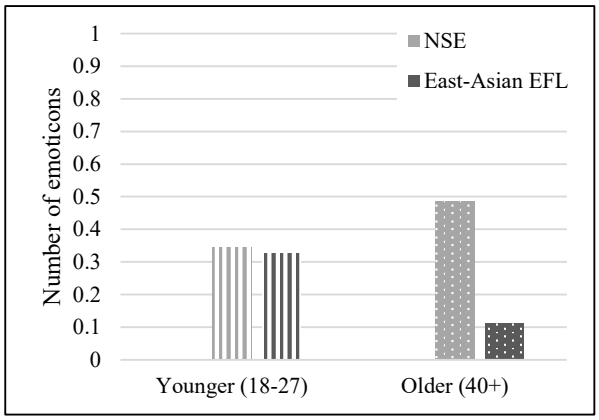


Figure 7: Emoticons by Age group by Native language

The fact that no interaction effect between Condition and Native language was found indicates that native speakers of English and East-Asian speakers of English as a foreign language accommodate to the same degree to their interlocutors' textism, emoji, and emoticon use in their English CMC. This suggests that any potential differences between NSE participants and East-Asian EFL participants in, for example, familiarity with English texting abbreviations, emoji, emoticons, or kaomoji do not result in more or less accommodation.

5. Conclusion

The present study analysed a corpus of experimentally elicited chat messages. Evidence was found of language style accommodation with textual and visual online language style elements, namely textisms, emoji, and emoticons. Participants were shown to converge their online language style to that of their interlocutors. This confirms Communication Accommodation Theory in a computer-mediated setting. In addition, age group and native language were found to affect people's use of textisms and emoticons, but did not affect accommodation.

6. Acknowledgements

I would like to express a big thanks to my two wonderful student assistants, Thị Thảo Anh Đặng and Noa ter Braak, for their help in developing the experimental stimuli and coding the corpus.

7. References

Adams, A., Miles, J. (2023). Examining textism convergence in mediated interactions. *Language Sciences*, 99, 101568.

Adams, A., Miles, J., Dunbar, N., and Giles, H. (2018). Communication accommodation in text messages: Exploring liking, power, and sex as predictors of textisms. *Journal of Social Psychology*, 158(4), pp. 474--490.

Brinberg, M., Ram, N. (2021). Do new romantic couples use more similar language over time? Evidence from intensive longitudinal text messages. *Journal of Communication*, 71(3), pp. 454--477.

Danescu-Niculescu-Mizil, C., Gamon, M., and Dumais, S. (2011). Mark my words! Linguistic style accommodation in social media. In *Proceedings of the 20th International Conference on World Wide Web*, pp. 745--754.

Evans, V. (2017). *The Emoji Code: How Smiley Faces, Love Hearts and Thumbs Up Are Changing the Way We Communicate*. Michael O'Mara.

Giannoulis, E., Wilde, L.R.A. (Eds.) (2019). *Emoticons, Kaomoji, and Emoji: The Transformation of Communication in the Digital Age*. Routledge.

Giles, H. (Ed.) (2016). *Communication Accommodation Theory: Negotiating Personal Relationships and Social Identities Across Contexts*. Cambridge UP.

Giles, H., Coupland, J., and Coupland, N. (Eds.) (1991). *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge UP.

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, 738278.

Hilte, L., Daelemans, W., and Vandekerckhove, R. (2024). Communicating across educational boundaries: Accommodation patterns in adolescents' online interactions. *Applied Linguistics Review*, 15(1), pp. 1--29.

Hilte, L., Vandekerckhove, R., and Daelemans, W. (2022). Linguistic accommodation in teenagers' social media writing: Convergence patterns in mixed-gender conversations. *Journal of Quantitative Linguistics*, 29(2), pp. 241--268.

Kroll, T., Braun, L.M., and Stieglitz, S. (2018). Accommodated emoji usage: Influence of hierarchy on the adaption of pictogram usage in instant messaging. In *Proceedings of Australasian Conference on Information Systems*.

Lee, J.S., Dressman, M. (2018). When IDLE hands make an English workshop: Informal digital learning of English and language proficiency. *Tesol Quarterly*, 52(2), pp. 435--445.

Lee, J.S., Sylvén, L.K. (2021). The role of Informal Digital Learning of English in Korean and Swedish EFL learners' communication behaviour. *British Journal of Educational Technology*, 52(3), pp. 1279--1296.

Lu, Y., Kroon, S. (2024). Elder Biaoqing: Investigating the indexicalities of memes on Chinese social media. *Chinese Semiotic Studies*, 20(1), pp. 71--93.

Markman, K.M., Oshima, S. (2007). Pragmatic play? Some possible functions of English emoticons and Japanese kaomoji in computer-mediated discourse. In *Association of Internet Researchers Annual Conference 8.0*, pp. 1--19.

Marko, K. (2022). "Depends on who I'm writing to": The influence of addressees and personality traits on the use of emoji and emoticons, and related implications for forensic authorship analysis. *Frontiers in Communication*, 7, 840646.

Muir, K., Joinson, A., Cotterill, R., and Dewdney, N. (2017). Linguistic style accommodation shapes impression formation and rapport in computer-mediated communication. *Journal of Language and Social Psychology*, 36(5), pp. 525--548.

Pickering, M.J., Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4, pp. 203--228.

Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9(5), pp. 1--6.

Siebenhaar, B. (2018). Accommodation in WhatsApp communication. In *Proceedings of the 6th Conference on Computer-Mediated Communication and Social Media Corpora*, p. 3.

Thurlow, C., Poff, M. (2013). Text messaging. In S.C. Herring, D. Stein, & T. Virtanen (Eds.), *Handbook of the Pragmatics of CMC*. Mouton de Gruyter, pp. 163--190.

Verheijen, L. (2017). WhatsApp with social media slang? Youth language use in Dutch written computer-mediated communication. In D. Fišer & M. Beißwenger (Eds.), *Investigating Computer-Mediated Communication: Corpus-Based Approaches to Language in the Digital World*. Ljubljana UP, pp. 72--101.

Wagner, T., Punyanunt-Carter, N., and McCarthy, E. (2022). Rules, reciprocity, and emojis: An exploratory study on flirtatious texting with romantic partners. *Southern Communication Journal*, 87(5), pp. 461--475.

Wang, H.C., Fussell, S.R., and Setlock, L.D. (2009). Cultural difference and adaptation of communication styles in computer-mediated group brainstorming. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, pp. 669--678.

Zeoob (2024). https://zeoob.com/generate-whatsapp-chat/

Zhang, Y., Liu, G. (2022). Revisiting informal digital learning of English (IDLE): A structural equation modeling approach in a university EFL context. *Computer Assisted Language Learning*, pp. 1--33.

8. Appendix

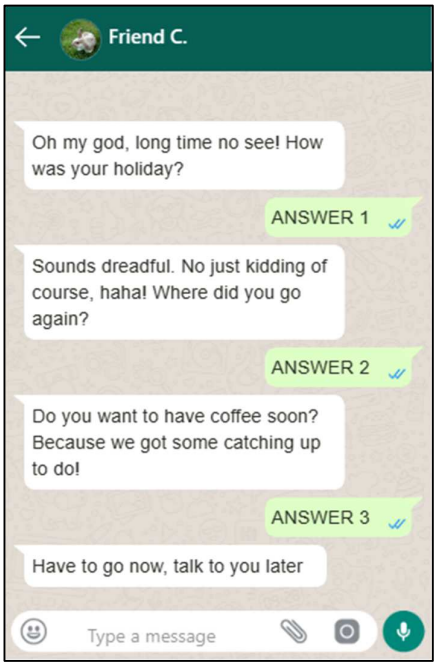


Figure 8: Chat conversation C., control condition

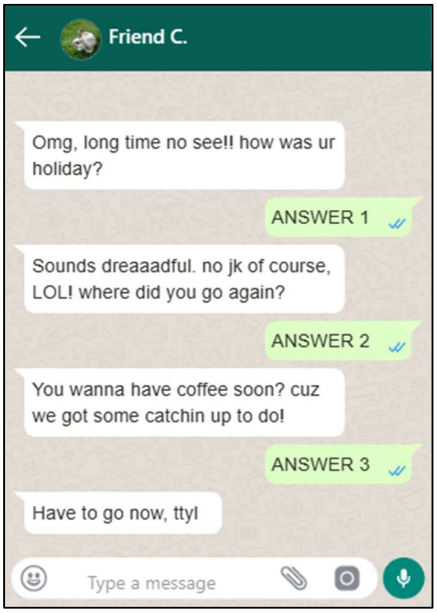


Figure 9: Chat conversation C., textisms condition

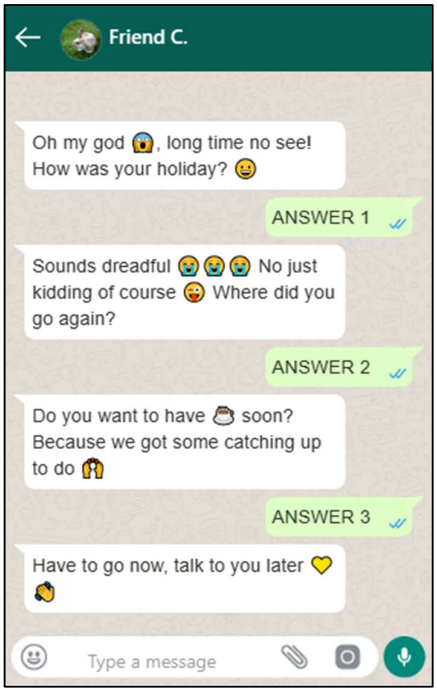


Figure 10: Chat conversation C., emoji condition

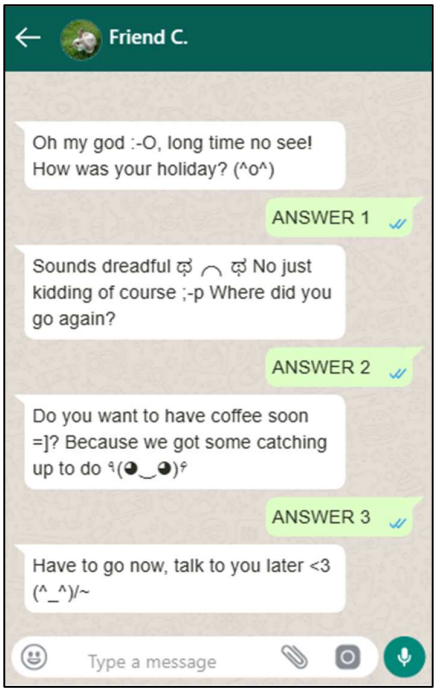


Figure 11: Chat conversation C., emoticons condition

POSTERS

Contested Landscapes: scripts as graphic and semiotic tools to social meaning

May Ahmar
Columbia University
E-mail: ma2550@columbia.edu

Abstract

This paper explores the linguistic and metalinguistic practices that activists on SNS applied to posts, lexicon and hashtags, in order to circumvent the censorship imposed on them on these social platforms during the Palestine-Israel war in 2021, and 2023-4. Activists posting about the Palestine-Israel war faced restrictions of posts by major SNS, removal of content and posts related to the war, as well as blocking users and deleting their accounts for posting specific words. As a result, activists and platform users devised a number of strategies to avoid being blocked, and keep the access to their content available. Among these practices, this paper focuses on script-fusing, script-disconnecting, respelling, misspelling, and abbreviations as graphic and semiotic tools employed by activists to help them keep access to their content on Facebook and Instagram.

The data are situated within the framework of graphic variation and graphic ideologies (Spitzmüller, 2015), as well as explored through a social semiotics' approach to these practices, which happen within a landscape that creates and reflect social meaning (Sebba 2007 and 2009). By using Arabic graphemes within a Latin-script word, and vice versa, as well as using the other tools that this paper delves into, activists and social actors are reclaiming power over language, lexicon and the social space they are interacting in, namely SNS, and the sub-spaces created within to share information, news and opinions about the topic at hand. This paper explores the graphic ideologies that exist within the linguistic landscape, that help in the creation of social meaning associated with these practices and their users. The practices reflect differences in social value and support an ideology that breaks the hierarchy imposed by the "powerful" SNS that control internet speech and freedom of expression, to subverting these powers and reclaiming power within the same space.

Keywords: script, orthography, semiotics

References

Cutler, C., Ahmar, M., & Bahri, S. (Eds.). (2022). *Digital orality: Vernacular writing in online spaces*. Springer Nature.

Jaffe, A., Androutsopoulos, J., Sebba, M., & Johnson, S. (Eds.). (2012). *Orthography as social action*. Walter de Gruyter, Incorporated.

Nuessel, F. (2015). Deviant orthography. *International handbook of semiotics*, pp. 291-301.

Panović, I. (2018). 'You don't have enough letters to make this noise': Arabic speakers' creative engagements with the Roman script. *Language Sciences*, *65*, pp. 70-81.

Sebba, M. (2007). *Spelling and society: The culture and politics of orthography around the world*. Cambridge University Press.

Spitzmüller, J. (2015). Graphic variation and graphic ideologies: A metapragmatic approach. Social semiotics, 25(2), pp. 126-141.

“Are There Any ‘Must Attend’ Lectures?”: Initial results from a Reddit corpus of cross-UK university student discussions

Marc Alexander
University of Glasgow
E-mail: marc.alexander@glasgow.ac.uk

Abstract

The aggregator and discussion site Reddit is structured around 'subreddits' on a given theme, and there are over fifty of these dedicated to student discussion of UK universities. As these are informal fora, separate from official institutional ones, they provide a relatively unfiltered insight into the makeup and preoccupations of students at these universities in recent years.

The corpus contains both Reddit posts and comments (anonymised but alongside associated metadata) from February 2011 to December 2019 (pre-pandemic) for 49 UK universities. (While 50+ UK universities have dedicated subreddits, some were too small or inactive to include in this study.) The universities included span a wide range of UK institutions, from ancient to newer and from institutions with a specialist focus to those more broad-based. In total, there are 8,275 posts and 59,602 comments in the corpus, totalling 2,898,249 words.

There are two areas into which the poster will present initial research. Firstly, the distinctive individual terms used at each institution, derived from a keyness analysis of each institution against the rest, processed in R, shows the preoccupations and local terminology in use by each institution. Secondly, and more importantly, by taking the corpus as a whole and analysing keyness against a general corpus alongside a study of n-grams, the areas students most ask questions about will be examined, to discover what this tells us about the ‘hidden curriculum’; those unwritten lessons about institutions which encode the assumptions and expectations university students are intended to acquire but are rarely explicitly taught. By examining the most frequent and distinctive terms and queries in this corpus, we can find information about the types of things students ask each other about institutions – and work out what universities should be working to make explicit and so reduce our unwritten expectations of students.

Keywords: discourse analysis, university students, keyness

Writing oral languages online: Ettounsi and Tamazight challenging standard ideologies

Soubeika Bahri
Independent scholar
E-mail: wafafedy31@gmail.com

Abstract

In post-revolutionary Tunisia, the use of Social Networking Sites and computer-mediated communication has led to the development of new literacy practices. These practices are characterized by the use of oral language, Ettounsi, in writing, with the medium of three types of scripts: Latin, Arabic, and Arabizi. The post-revolutionary period is also marked by the re-emergence of the indigenous language Tamazight, which can be written in three different scripts: Tifinagh, Latin, and Arabic. Using a mixed method design combining a corpus of Facebook posts (N=1043 posts) collected from twelve pages and 19 semi-structured interviews and content analysis method, the study argues that these new literacy practices in the context of Tunisia are used to communicate ideologies.

Keywords: Ettounsi, script practices, Tamazight

1. Introduction

In post-revolutionary Tunisia, the use of Social Networking Sites and computer-mediated communication has led to the development of new literacy practices. These practices involve using spoken language, Ettounsi, in written form through three types of scripts: Latin, Arabic, and Arabizi (a combination of Latin letters and numbers) (Alghamdi and Petraki 2018; Heghegh 2021; Cutler et al., 2022). This new practice has presented challenges to conventional concepts of mother tongue, official, national, and standard language, which are typically associated with Arabic language in the context of Tunisia. Furthermore, the post-revolutionary period has seen the re-emergence of the indigenous language Tamazight, which can be written in three different scripts: Tifinagh (the Lybico-Berber alphabet used to write Amazigh languages), Latin, and Arabic. The appearance of Tamazight has complicated the controversy over the notion of the mother tongue.

This study utilizes the methods of discourse-centered online ethnography (Androutsopoulos 2007) and computer-mediated discourse analysis (Herring 2004), focusing on the online communities of Facebook to investigate the emerging ideological debate between two groups of Tunisians: those who believe Ettounsi should be standardized and those who think it should remain an oral language. The discussion extends to encompass an arising debate over which language should be considered the mother tongue. This includes those who think Tamazight, as the language of the land, is the mother tongue, and pan-Arabism/nationalists who think Standard Arabic is the only mother tongue while Ettounsi is nothing but a spoken vernacular. The implications of these competing ideologies will be discussed according to the sociolinguistic reality of Tunisia today.

2. Script variations in literature

Previous research on non-standard writing has mainly focused on Latin-based orthographies for non-Latin alphabetized languages, a process known as Latinization. This has been studied in various languages, including Haitian (Schieffelin and Doucet 1994), Cantonese (Lee 2007) and Arabic (Palfrreyman & Khalil 2007). Digital non-standard writing also includes alternative orthographic practices. For instance, some languages maintain their standard scripts while losing secondary orthographic markings, such as diacritics. Additionally, Arabic vernaculars use both the standard Arabic alphabet and the emerging Arabizi system, which uses non-standard Roman letters to write Arabic scripts, along with Roman letters and numerals.

Campaigns against the use of Arabizi, a form of westernization, were led by Arabic enthusiasts. They believed Arabizi to be a "deviant" code and expressed concerns about its impact on the Arabic language. Muhammed et al. (2011) reported that those who reject Arabizi do so out of fear of its effects on the younger generation, leading them to embrace a more westernized identity and gradually lose their Arabic identity. The use of Arabizi is also often seen as a threat to the language of the Quran (Muhammed et al., 2011), causing concerns regarding the potential loss of the future generations' Islamic identity. The use of Latin script or Arabizi in writing Arabic has been well-documented, but Tamazight presents a different case. The choice of writing systems has sparked extensive debate in the context of Morocco and Algeria. Souag (2004) argues that the Kabyle movement in Algeria strongly favors the Latin script, while in Morocco, Souleimani (2013) indicates that Neo-Tifinagh was officially chosen as the primary script for Tamazight and received significant governmental support for its inclusion in school curricula. However, in both countries, activists, intellectuals, and the general population continue to have disagreements regarding the script choices made by the authorities. Identity, ideology, and linguistic distinctions play crucial roles in driving these script choices.

3. Method

Based on the fact that Facebook is the most used social media platform in Tunisia (Statista, 2021), this article utilized a mixed method design. It combined a corpus of 1043 Facebook screenshots that combined posts and comments. The data was collected from twelve Tunisian Facebook pages selected randomly and 19 semi-structured interviews conducted from 2020 through 2023 to investigate the attitudes and motives behind the scriptural variations in writing Ettounsi and Tamazight. The Facebook data involved Tamazight posts (N=462) written in Tifinagh, Latin, and /or Arabic script. The rest of the posts were in Ettounsi (N=591) written in Arabic, Latin, or Arabizi script. Table 1 shows the number of Facebook posts by script for each language.

Tamazight			Ettounsi		
Tifinagh	Latin	Arabic	Arabic	Latin	Arabizi
125	246	91	203	137	251

Table 1: Facebook posts by language and script

The study used content analysis to examine and quantify the relationship between topic types and each of the scripts in use. The data was initially organized and categorized into 10 themes. After identifying the initial themes, the researchers reviewed and refined the data by regrouping related themes and combining categories to create a new list of three general themes for the use of the varied scripts to write Tamazight and three others to write Ettounsi. However, some content remained uncategorized due to lack of relevance or uncertainty. As a result, some posts and comments were classified as “miscellaneous”.

4. Writing Ettounsi in varied scripts

The current sociolinguistic profile of Tunisia no longer revolves around just two dominant languages (Standard Arabic and French). The Tunisian identity is no longer seen as static, uniform, and exclusively "Arab". This change is evident through the emergence of new and diverse literacy practices in literature, street signs, advertisements, and most notably on social media, particularly Facebook. The representation of each script in digitally-mediated discourse can be seen as a form of social action (Sebba 2009) or an attempt to redefine the linguistic hierarchy in Tunisia, shaped by competing ideologies for power and dominance.

4.1 Writing Ettounsi in Arabic script

The colloquial variety of Arabic known as Ettounsi, which was traditionally spoken, is now being used in written form on digital platforms. When Ettounsi is written in Arabic script on Facebook, it showcases distinct lexical and stylistic characteristics. Many users adhere to the spelling rules of Standard Arabic when representing Ettounsi on Facebook.

Others, however, prefer a pronunciation-based approach over a strict Standard Arabic spelling. The relaxed spelling incorporates phonological and morphological features. This suggests that orthography on social media is no longer exclusively controlled by policymakers or institutions. Subsequently, the status of the Ettounsi language has undergone significant changes in the past decade. This shift has led to proposals for the official use of Ettounsi in education, particularly by the new association called Derja. This change is significant and indicative of a shift. It's worth noting that the same association has initiated a project to standardize Ettounsi despite the ongoing freedom and creativity that users maintain in their practices of writing the language. The choice to write Ettounsi in Arabic script is motivated by the desire to communicate with a wider audience who may not be comfortable reading Standard Arabic or French. There's also a political aspect to it, as it's seen as a way to convey ideas and opinions with spontaneity and authenticity, particularly on topics like religion and politics following the 2011 revolution.

4.2 Writing Ettounsi in Latin and Arabizi script

A significant number of interactions on Ettounsi Facebook are written in Latin script or Arabizi. Familiarity with this system and personal preferences are the main factors influencing its use. Several studies have found that Arabizi serves as an indicator of trendiness, coolness, and being up-to-date, especially among the youth (Muhammed et al. 2011). Therefore, Arabizi is a noticeable sociolinguistic marker of youth culture on Tunisian Facebook. The use of Arabizi, Latin, or a combination of both styles in writing is no longer simply a matter of mapping the language's sound system to a set of conventions; it appears to be a discursive practice. Writing in Ettounsi using the Latin alphabet or Arabizi is often associated with informality and tolerance. It typically occurs to discuss social topics such as family, entertainment, personal concerns, and other miscellaneous subjects. The use of Ettounsi in different types of writing and various contexts reinforces its connection with ease of expression and creativity, as well as its ability to accommodate all the phonemes of Ettounsi. When written in Romanized and Arabizi scripts, Ettounsi can be seen as a move away from the literary norms of Standard Arabic, which are challenging to master without formal education.

5. Writing Tamazight in varied scripts

The use of Ettounsi on Facebook is growing among Tunisians who identify as Imazighen (singular: Amazigh). These individuals utilize various language scripts, each of which is influenced by interactive functions, social values, meanings, and ideologies. Over the past decade since the 2011 Revolution, Arabic, Latin, and Tifinagh have emerged as three distinct writing

systems visible on Tunisian Amazigh Facebook.

5.1 Writing Tamazight in Arabic

Tunisian Amazigh people tend to use Arabic script to write Tamazight more often in religious contexts than in other forms of cultural expression. For example, they use Arabic to exchange greetings and wishes during Eid (Muslim feast) and the Muslim New Year. The use of Arabic script in similar contexts serves an affective function through which Imazighen align themselves with the religious meaning often associated with Arabic. Imazighen Facebook users often find themselves compelled to engage in discussions about their spiritual identity in a setting where Arabic is emphasized as the language of God and considered superior to the indigenous Tamazight. The use of Arabic script in religious contexts is motivated by the need to emphasize Imazighen's religiosity. According to many Imazighen, it is a crucial discursive strategy to counter the Islamic discourse where Tamazight is considered useless and secular, or even profane, representing a pagan past that the Imazighen should dissociate themselves from in order to be true Muslims (El-Masud, 2014).

5.2 Writing Tamazight in Tifinagh

It is important to remember that using Tifinagh in logos, tattoos, or mottos helps to promote Tamazight through its unique writing system. This also challenges the idea that Tamazight is solely a spoken language, allowing those who identify with Tamazight to show their role as integral members of the Amazigh community (Davis 2013). Presenting Tamazight as a language with its own distinct writing system and emphasizing that Tifinagh is compatible with modern keyboard technology responds to the demands of modernity and enhances the social and cultural status of Tamazight. However, it's important to note that Tifinagh remains limited in its impact unless accompanied by a Latin or Romanized script and a translation, as very few people are able to read it.

5.3 Writing Tamazight in Latin

Tunisian Amazigh people use Latin orthography for writing cultural materials, events, poems, translations, and other forms of literary works. The Latin script is considered a symbol of power, progress, and modernity. In the context of Tamazight, it is argued that the Latin script is the default based on interactions on the Tunisian Imazighen Facebook data. The choice of a Latin-based writing system for Tamazight is connected to beliefs about the importance of French language and culture. The French orthography is commonly seen as a practical and modern system for writing Tamazight in academic and activist circles (Souag, 2004). Additionally, French serves as a meta-communicative tool to aid in learning Tamazight.

6. Scripts in ideological conflict

The Tunisian situation appears to be similar to the one in Morocco, where the use of spoken vernacular in writing has started to disrupt the traditional language hierarchy. As mentioned before, there is a growing presence of writing in Ettounsi in digital media and recent publications, and efforts are being made by the Derija Association to promote Ettounsi literacy and standardize the language as the mother tongue of most Tunisians. However, it's evident from discussions on Facebook that for Tunisians who identify as Imazighen, the notion of mother tongue refers to Tamazight rather than Ettounsi. One interviewee expressed, "Tamazight is my identity, language, freedom, and life. I am Amazigh. I use the Arabic language to read the Quran and to pray. God said to be Muslim. He did not say to be Arab. Ettounsi is to communicate; it is a dialect, not a language." "The Tamazight language is considered the mother tongue of those born in North Africa. This statement brings attention to the competing ideologies that have emerged in Tunisian sociopolitical circumstances during the digital democratization. Imazighen perceive Ettounsi as primarily an oral language, even as it acquires written literary value. On the other hand, Tamazight is seen as the sole mother tongue of North Africans, as it has a historically established writing system known as Tifinagh, even though its practical significance tends to be more symbolic. Further research is needed in this area, particularly as sociopolitical changes and the growth of the digital space have given rise to multilingual and multi-literacy practices within the linguistic repertoire. This has sparked new forms of discrimination and broader debates about what constitutes a written language, which scripts should be used for what purpose, and what is considered a "mother tongue".

7. References

Androutsopoulos, J. 2007. Language choice and code-switching in German-based diasporic web forums. In B. Danet, & S. Herring (Eds.), The multilingual internet. Language, culture and communication online (pp. 340--361). Oxford: Oxford University Press.

Allehaiby, Wid. H, 2013. Arabizi: an analysis of the Romanization of the Arabic Script from a Sociolinguistic perspective. Arabi World English Journal. Vol. 3 (3).

Al-Khalil, M., and Palfreyman, D. 2003. “A funky language for teenzz to use”: Representing Gulf Arabic in instant messaging. *Journal of Computer-Mediated Communication*, 9 (1).

Davis, J. 2013. Learning to Talk Indian: Ethnolinguistic identity and language revitalization in the Chickasaw renaissance. Ph.D. Thesis. University of Colorado, Boulder.

4:497-513.

Sebba, Mark. 2009. *Spelling and society: The culture and politics of orthography around the world*. Cambridge: Cambridge University Press.

Souag, L. 2004. "Writing Berber Languages: a quick summary". Retrieved from https://web.archive.org/web/20041205195808/www.geocities.com/lameens/Tifinagh/index.ml, June 1, 2024.

Politicizing public health: The discourses around public health organizations

Tatiana Schmitz de Almeida Lopes, Fernanda Peixoto Coelho, Renata Sant'Anna Lamberti Spagnuolo

Pontifícia Universidade Católica de São Paulo - Faculdade de Tecnologia da Praia Grande

E-mail: profatatischmitz@gmail.com, fc.o2@me.com, reslamberti@gmail.com

Abstract

Before the pandemic, public health organizations and disease control agencies like the WHO, CDC, ANVISA, ECDC, and PAHO were relatively unknown to the general public, often operating in the background of healthcare and health policy systems. However, with the onset of the COVID-19 pandemic, these organizations quickly became household names. The pandemic has heightened public awareness of these agencies globally, especially in countries or regions severely impacted by COVID-19 (such as the US and Brazil). Concurrently, the politicization of the pandemic led to significant criticism of these agencies regarding their management strategies. Measures such as face mask mandates, vaccination policies, social distancing, lockdowns, and other actions enforced by local officials and supported by health authorities became points of ideological contention among groups with varying political leanings. This led to the creation of conflicting representations that framed these organizations in various ways, supported by a myriad of discourses based on a range of ideologies. This clash of ideologies gave rise to both positive and negative representations. To date, no study exists that seeks to provide a comprehensive view of the representations of health organizations on social media. To fill this gap, the goal of the current study is to carry out a corpus-based analysis of the discourses surrounding health organizations in Brazilian tweets in Portuguese. A corpus of ca. 88K tweets was collected whose messages included mention of the major world health organizations. The current corpus was collected using the Python tool "Twarc". We ran searches using different search terms, such as "pandemic," "WHO", "PAHO", "ANVISA, "CDC", "ECDC", "Coronavirus" and "Covid-19" and saved the output from all of these searches to a single JSON (JavaScript Object Notation) file, which contained the text posted and its related metadata in a structured format. Specialized scripts were developed to clean up the JSON file, remove duplicates, and retain only the relevant text and metadata information. The FAIR Data principles intend to be respected since the research is financed by the "CNPQ" (National Council for Scientific and Technological Development) and we intend to make the data collected available on the Open Science Portal for future research. The analysis employed Lexical Multidimensional Analysis (Berber Sardinha, 2019, 2021; Berber Sardinha & Fitzsimmons-Dooley, 2024; Fitzsimmons-Doolan, 2019, 2023), an extension of Multidimensional Analysis (Biber, 1988), based on the identification of factors based on shared lexical patterns, precisely unlike a regular multidimensional analysis, which is concerned with the functional description of register variation. A lexical multidimensional analysis is centered on the description of discourses, ideologies, themes, and other lexis-based constructs in corpora (Berber Sardinha, 2017, 2021, 2023; Berber Sardinha & Fitzsimmons-Doolan, in prep.; Clarke et al., 2021, 2022; Fitzsimmons-Doolan, 2014, 2019, 2023). These factors, which were interpreted as dimensions that encapsulate the discourses shaping the representations of health agencies during the COVID-19 crisis, were interpreted and will be detailed in the presentation. Like this, the choice of lexical AMD aims to identify understandings, social perceptions, public trust and denialism in communications coming from such agencies, promoting better forms of communication coming from such agencies, avoiding misunderstandings regarding health guidelines, since such guidelines affect society and democracy in defense of the protection of life as a fundamental right guaranteed in the Universal Declaration of Human Rights.

Keywords: Multidimensional Lexical Analysis, Social Media, Infodemic, Pandemic, Health Agencies

References

Berber Sardinha, T. (2019). Using multi-dimensional analysis to detect representations of national culture. In T. Berber Sardinha & M. Veirano Pinto (Eds.), Multi-dimensional Analysis: Research Methods and Current Issues (pp. 231-258). London: Bloomsbury.

Berber Sardinha, T. (2021). Discourse of academia from a multi-dimensional perspective. In E. Friginal & J. Hardy (Eds.), The Routledge Handbook of Corpus Approaches to Discourse Analysis (pp. 298-318). Abingdon: Routledge.

Berber Sardinha, T., & Fitzsimmons-Doolan, S. (2024). Lexical Multidimensional Analysis. Cambridge: Cambridge University Press.

Biber, D. (1988). Variation Across Speech and Writing. Cambridge: Cambridge University Press.

Fitzsimmons-Doolan, S. (2019). Language ideologies of institutional language policy: Exploring variability by language policy register. Language Policy, 18(2), 169-189.

Fitzsimmons-Doolan, S. (2023). 21st century ideological discourses about US migrant education that transcend registers. Corpora, 18(2), 143-173.

Harnessing Twitter Corpus for Neural Machine Translation in Low-Resource Languages: A Case Study of Spanish-Galician

María do Campo Bayón
E-mail: maria.docampo@autonoma.cat

Abstract

This paper presents the development of a Neural Machine Translation (NMT) engine tailored for social media, specifically focusing on translating from Spanish to Galician within a low-resourced language context. By utilizing a specialized Galician Twitter corpus, we demonstrate the methodologies involved in corpus construction and training. Evaluation methods included both automated metrics and non-inferiority analysis, complemented by human assessment. Our findings underscore the importance of expanding corpus size and employing back-translation to improve performance. Notably, we observed that high-quality language in science popularization accounts significantly enhanced translation accuracy. Professional linguists also identified opportunities for post-editing to further refine machine-translated content. This study highlights Twitter's potential as a valuable source of monolingual data, extending beyond sentiment analysis. Moreover, it offers a replicable framework for promoting linguistic diversity in digital environments, especially for low-resourced languages. Through social media, our aim is to actively engage younger Galician speakers, addressing the decline in speaker numbers.

Keywords: Twitter corpus, monolingual corpus, low-resourced languages, NMT, Galician corpus.

1. Introduction

The rapid growth of social media has highlighted the need for effective translation systems capable of handling the unique characteristics of user-generated content. This paper presents the development of a Neural Machine Translation (NMT) engine specifically designed to translate content from Spanish to Galician, focusing on the challenges associated with low-resourced languages. Using a specialized Galician monolingual Twitter corpus, the study explores the corpus construction process, the training methodologies employed, and the evaluation of the engine's performance.

2. Context and Need for the Study

Galician, a minority language spoken in the region of Galicia in Spain, has been experiencing a decline in speakers, especially among the younger population (IGE, 2019). Developing an NMT system for Galician is essential to bolster its presence on digital platforms and engage younger demographics. While a few Galician-Spanish NMT systems exist (for example, Ortega *et al*., 2022), they are not tailored for the social media domain, which has unique linguistic characteristics and informal styles not adequately handled by general-domain MT systems. Hence, there is a specific need for an NMT system focused on social media content to ensure high-quality translations that reflect the natural use of language in these settings.

3. NMT training and corpus construction

We chose a deep neural network architecture, specifically the transformer model, for our NMT engine (do Campo Bayón, 2023). Open-source platforms such as JoeyNMT (Kreutzer *et al.*, 2019) and Marian (Junczys-Dowmunt *et al.,* 2018) were utilized to ensure accessibility and reproducibility.

3.1 Corpus Compilation

Given the scarcity of Galician data and the absence of social media-specific corpora, we combined general and specific data sources. Initial data collection in 2020 leveraged existing bilingual resources like the Paracrawl corpus (Bañón *et al.*, 2020). The Paracrawl corpus, published by the European Commission, offered 1,879,651 bilingual segments, providing a substantial general-domain dataset. To address the lack of social media-specific data, we followed a similar approach to Lohar *et al.*, 2019 and created a monolingual Galician Twitter corpus. This involved selecting tweets from official Galician institutional accounts, ensuring the quality of the language in these accounts in terms of grammar, spelling, and naturalness, but still reflecting the informal and conversational nature of social media texts. Python's *snscrape* library facilitated the extraction and conversion of tweets into a usable format, followed by thorough cleaning and filtering using the MTUOC library (Oliver, 2020).

3.2 Ethical Considerations and Data Usage

The data collection process adhered to ethical guidelines and Twitter's data usage policies. We ensured that the collected data were publicly accessible tweets, and no private data were used. Researchers registered as academic users of the Twitter API to legally collect and analyse the data. Furthermore, we anonymized the data to protect user privacy and complied with all applicable regulations and guidelines.

3.3 Data Augmentation and Back-Translation

To maximize data utility and enhance the engine's performance, we employed various strategies, including back-translation. This process involved translating the Galician tweets into Spanish using Google Translate, creating a bilingual corpus for training. The combined use

of Paracrawl's general corpus and the specific Twitter corpus was critical to addressing the diversity of social media content. Back-translation was performed automatically to efficiently handle the large volume of data.

4. Training Methodologies

4.1 Training Process

The training process entailed meticulous parameter tuning, including batch size, learning rate, and loss function optimization. Techniques such as tokenization, truecasing, and the removal of non-relevant content were applied using Python's MTUOC library. The processed data were split into training, validation, and evaluation sets to ensure robust model performance. A first version of the engine was trained using JoeyNMT through MUTNMT[1] platform.

4.2 Iterative Refinement

A pilot test was conducted with this first version of the engine and was evaluated automatically and according to the principle of non-inferiority (do Campo Bayón and Sánchez Gijón, 2024). The BLEU score obtained was 70.63. Insights from the non-inferiority evaluation indicated the need for a larger Twitter-specific corpus to improve translation quality for short sentences (tweets formed by 2-10 words). Consequently, to enhance the performance of the engine, it was decided to re-train the model using the Marian technology and a larger sample of short segments. This new training incorporated additional Galician institutional Twitter accounts, expanding the corpus to 262,785 unique sentences. The refined training process utilized this expanded corpus alongside the Paracrawl data, significantly enhancing the model's performance.

5. Evaluation and Results

The NMT engine's performance was evaluated using automatic metrics (BLEU) alongside two human evaluations: DQF-MQM professional linguists classifying translation errors, providing qualitative insights; and the non-inferiority analysis comparing machine translations to Tweets originally written in Galician by Galician Twitter users.

The results demonstrated that increasing the size of the in-domain Twitter corpus and employing back-translation markedly improved the engine's performance. The BLEU score was 85%. Professional linguists noted that the machine translations required minimal post-editing, highlighting the engine's effectiveness, and Twitter users generally did not perceive our translations as inferior (do Campo Bayón, 2023).

6. Conclusion

This paper presents the development and evaluation of a Spanish-Galician NMT engine tailored for social media content. The findings highlight the effectiveness of augmenting the training corpus and utilizing back-translation techniques. The promising results obtained underscore the viability of social media as a data source for NMT training and the importance of supporting low-resourced languages in the digital age. By fostering engagement with younger demographics, such initiatives can contribute to the revitalization and preservation of linguistic diversity.

7. References

Bañón M., Chen, P., Haddow, B., Heafield, K., Hoang, H., Esplà-Gomis, M., Forcada, M., Kamran, A., Kirefu, F., Koehn, P., Ortiz Rojas, S., Pla Sempere, L., Ramírez-Sánchez, G., Sarrías, E., Strelec, M., Thompson, B., Waites, W., Wiggins, D. and Zaragoza, J.. (2020). ParaCrawl: Web-Scale Acquisition of Parallel Corpora. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp 4555--4567.

Do Campo Bayón, M. (2023). *Traducción automática neuronal para lenguas con recursos reducidos. Evaluación de los usuarios según el principio de no inferioridad.*Doctoral Thesis. UAB.

Do Campo Bayón, M. and Sánchez-Gijón, P. (2024). Evaluating NMT using the noninferiority principle. In *Natural Language Engineering.* Cambridge.

Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T., Hoang, H., Heafield, K., Neckermann, T., Seide, F., Germann, U., Fikri Aji, A., Bogoychev, N., Martins, A. F. T. y Birch, A. (2018). Marian: Fast Neural Machine Translation in C++. In *Proceedings of ACL 2018, System Demonstrations*, pp. 116-.121.

Kreutzer, J., Bastings, J. and Riezler, S. (2019) Joey {NMT}: A Minimalist {NMT} Toolkit for Novices. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations,* pp. 109—114.

Lohar, P., Popović, M., Alfi, H. and Way, A. (2019). A systematic comparison between SMT and NMT on translating user-generated content. In *20th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2019),* pp. 7—13.

Oliver, A. (2020). MTUOC: easy and free integration of NMT systems in professional translation environments. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pp. 467--468.

Ortega, J. E., de-Dios-Flores, I., Campos, J. R. P., and Gamallo, P. (2022). A neural machine translation system for Spanish to Galician through Portuguese transliteration. Demo presented at the *15th International Conference on Computational Processing of Portuguese (PROPOR 2022*

[1] Available in: https://ntradumatica.uab.cat/.

From the Web to the Street, and Back: A Semiotic Approach to the Circulation of Militant Writings

Claire DOQUET,
Université de Bordeaux, Lab-E3D

Chenyang ZHAO,
Université Sorbonne Nouvelle, Clesthia,
E-mail: claire.doquet@u-bordeaux.fr, chenyang.zhao@sorbonne-nouvelle.fr

Abstract

This proposal examines two discursive contexts originating from militant culture: Internet memes and feminist collages. These two modes of expression share the common feature of diversion and the ability to be recontextualised in different socio-discursive spheres (Marino, 2015). Both memes and feminist collages mobilise a form of serialisation, manifested in the repetition of content that lend themselves to endless reinvestment, with meanings that vary according to the context of the display, physical or digital.

We chose to analyze a dozen feminicide collages during the movement for the constitutionalization of abortion in July 2022, photographed and distributed on the internet (https://www.instagram.com/collages_feminicides_paris/), and a dozen memes created online and pasted or printed on protest supports (https://twitter.com/memes_de_manif). Using a qualitative and enunciative approach, and employing the categories established by Authier-Revuz (2020), we were able to analyze these creations as "discourses of others", represented in various semiotic materialities, which promotes the (re)production and circulation of internet discourses. The diversion through allusion (Dancygier et al., 2017) as a characteristic process of internet communication is thus reinforced by the change of the exhibition context. This material and media transposition also promotes the generation of new enunciative layers, superimposed in forms of links or hashtags, constantly creating a discursive network.

Keywords: Semiotics, Discourse of others, Internet memes actualization

References

Authier-Revuz J. (2020) *La Représentation du Discours Autre : principes pour une description.* Paris-Berlin : De Gruyter.

Collages Féminicides Paris, C. F. (2021) *Notre colère sur vos murs*. Denoël.

Dancygier, B. and Vandelanotte, L. (2017) Internet memes as multimodal constructions. *Cognitive Linguistics,* vol. 28, p.565-598. https://doi.org/10.1515/cog-2017-0074

Marino, G. (2015) Semiotics of spreadability: A systematic approach to Internet memes and virality. *Punctum*, 1(1): 43-66.

Bringing CMC corpora to the people: improving the usability of the French CoMeRe collection

Achille Falaise
Université Paris Cité, Laboratoire de Linguistique Formelle, CNRS
achille.falaise@cnrs.fr

Abstract

This poster aims to present an ongoing attempt to make CMC corpora more usable for linguists, starting with the CoMeRe corpora collection.

Keywords: CMC corpora, NLP processing of CMC texts, corpus linguistics

1. Context

Although CMC corpora have been readily available for some time (Frey et al., 2020), they are not used as much as they could be by linguists, at least for French. The reason may be that although these corpora are *available* (i.e. for download), they are not easy to use. Even when properly TEI-CMC encoded, they are tricky to work with using corpus linguistic tools, for many reasons, including :

- They are sometimes filled with meta-information (such as anonymisation traces, event notifications, etc.), which makes the actual linguistic content difficult to extract.
- They are not tokenised, which is not a trivial task.
- Parsing them is still a challenging task.

This poster aims to present an ongoing attempt to make CMC corpora more usable for linguists, starting with the CoMeRe collection. CoMeRe[1] (Chanier et al., 2014) is a French CMC collection consisting of 15 corpora (74 million words) covering several CMC genres:

- SMS (Ledegen 2010; Antoniadis, Chabert & Zampa, 2010; Panckhurst et al. 2016),
- wiki discussions (Poudat et al., 2014),
- tweets (Longhi, 2006; Longhi, 2013),
- weblogs (Abendroth-Timmer et al., 2009),
- emails (Reffay et al., 2009),
- forums (Reffay et al., 2009),
- text chat (Falaise, 2005; Yun & Chanier, 2011; Reffay et al., 2009),
- multimodal (Chanier et al., 2009; Chanier & Audras, 2011; Chanier & Wigham, 2011).

Our efforts will be focused on the 12 non-multimodal corpora.

2. Project

We will tokenise, POS-tag, lemmatise and syntactically parse these corpora, using an existing software processing pipeline[2] that supports SpaCy and Stanza, and is able to preserve the XML annotations of the corpora. The lemmatisation layer will enable searches to be carried out using a normalised orthography.

We will use existing general purpose Universal Dependencies models for French and evaluate the quality of tokenisation using the French Social Media Bank (Seddah et al., 2012). As there is no reference corpus for CMC UD parsing, we will only perform a minimal quality assessment for POS tags and lemmas against a small ad hoc validation corpus.

The resulting collection will be formatted in XML/CONLL-U and TXM format (Heiden, 2010), and will be made available for search in ScienQuest (Falaise, Tutin & Kraif, 2011). We intend to perform some ad hoc post-processing of the data per corpus and per format (creating meta-tags for tags that are often confusing for parsers, formatting/presenting metadata in a convenient way, etc.) in order to make it more convenient to use in corpus linguistic tools.

3. References

Abendroth-Timmer, D., Bechtel, M., Chanier T. & Ciekanski, M. (2010). (dir.) *LETEC (Learning and Teaching Corpus) Infral*. Mulce.org : Clermont Université.

Antoniadis G., Chabert G. & Zampa V. (2011). Alpes4science: Constitution d'un corpus de SMS réels en France métropolitaine. *Colloque TEXTOS: dimensions culturelles, linguistiques et pragmatiques*. Congrès annuel de l'ACFAS, Sherbrooke, Canada.

Chanier, T., Reffay, C., Betbeder, M-L., Ciekanski, M. & Lamy, M-N. (2009). *LETEC (Learning and Teaching Corpus) Copéas*. Mulce.org : Clermont Université.

Chanier, T. & Audras, I. (2011). (editors). *LETEC (Learning and Teaching Corpus) Tridem 2006*. Mulce.org : Clermont Université.

Chanier, T. & Wigham, C.R. (2011). (editors). *Learning and Teaching Corpus (LETEC) of ARCHI21*. Mulce.org : Clermont Université.

Chanier, T., Poudat, C., Sagot, B., Antoniadis, G., Wigham, C. R., Hriba, L., Longhi, J. & Seddah, D. (2014). The CoMeRe corpus for French: structuring and annotating heterogeneous CMC genres. Special issue on Building And Annotating Corpora Of Computer-Mediated Discourse: Issues and Challenges at the Interface of Corpus and Computational Linguistics. *Journal of Language*

[1] https://hdl.handle.net/11403/comere

[2] https://gitlab.com/llf-pli/corpus-pipeline

Technology and Computational Linguistics. p. 1-31. [http://www.jlcl.org/2014_Heft2/Heft2-2014.pdf]

Falaise, A. (2005). Constitution d'un corpus de français tchaté. Actes de RECITAL 2005, Dourdan. [https://hal.science/hal-00909667]

Falaise A., Tutin, A., Kraif O. (2011). Une interface pour l'exploitation de corpus arborés par des non informaticiens : la plate-forme ScienQuest du projet Scientext. *Traitement automatique des langues*, 52-3, p. 103-128. [https://www.atala.org/content/tal_52_3_4]

Frey, J.-C., König, A., Stemle, E., Falaise, A., Fišer, D. & Lüngen, H. (2020). The FAIR Index of CMC Corpora, *in* Julien Longhi, Claudia Marinica (eds.), *CMC Corpora through the prism of Digital Humanities, Collection Humanités numériques*, Éditions l'Harmattan.

Heiden, S. (2010). The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme. In 24th Pacific Asia Conference on Language, Information and Computation (p. 10). Sendai, Japan. [https://shs.hal.science/halshs-00549764]

Ledegen, G. (2010). Contact de langues à La Réunion : «On ne débouche pas des cadeaux. Ben i fé qoué alors ?». *Langues et Cité, Langues en contact*, 16, 9-10. [http://hal.archives-ouvertes.fr/hal-00879323]

Longhi, J. (2006). De intermittent du spectacle à inter-mittent : de la représentation à la nomination d'un objet du discours, *Corela*, 4-2. [http://corela.revues.org/457]

Longhi, J. (2013). Essai de caractérisation du tweet politique, *L'Information grammaticale*, n°136, p.25-32.

Panckhurst R., Détrie C., Lopez C., Moïse C., Roche M. & Verine B. (2016). *88milSMS. A corpus of authentic text messages in French*. ISLRN: 024-713-187-947-8.

Poudat, C, Grabar, N., Kun, J. & Chanier, T. (2014). Wikiconflits, un corpus extrait de Wikipédia : principe et méthode d'élaboration. In Poudat,C., Grabar , N. Kun, J., & Paloque-Berges, C. (2015). *Corpus Wikiconflits, conflits dans le Wikipédia francophone*.

Reffay, C., Chanier, T., Lamy, M.-N. & Betbeder, M.-L. (2009). (editors). *LETEC corpus Simuligne*. Mulce.org : Clermont Université.

Seddah, D., Sagot, B., Candito, M., Mouilleron, V. & Combet, V.. The French Social Media Bank: a Treebank of Noisy User Generated Content. *COLING 2012 - 24th International Conference on Computational Linguistics*, Kay, M. & Boitet, C. (eds.), Dec 2012, Mumbai, India. [https://hal.science/hal-00780895]

Yun, H. & Chanier, T. (2011). (editors). *LETEC corpus FAVI*. Mulce.org : Clermont Université.

Lexical Variation of the Albanian Language used in computer-mediated communication and the challenge for processing

Besim Kabashi
Friedrich-Alexander-Universität Erlangen-Nürnberg
Bismarckstraße 6, 91054 Erlangen, Germany
besim.kabashi@fau.de

Abstract

In addition to the standard variant of a language, a lot is also spoken and written in non-standard variants. The processing of data that is available in a non-standard variant is associated with many difficulties because the resources and tools were initially created and developed for standard variants and are either missing or insufficient and not good enough for non-standard variants. However, this data is very diverse and prove to be richer than the standard language data, i.e. is also important and must be taken into account and processed. Here we present our work dealing with the processing of lexical variants in the Albanian language used in computer mediated communication. In particular, we are concerned with the normalization of lexical variants and their tagging. We have collected texts from social media that are written in non-standard language, i.e. variants. We discuss them, the phenomena and the steps of processing.

Keywords: The Albanian language, Lexical Variation, Computer-mediated Communication, Normalisation, PoS-Tagging

1. Introduction

There is a lot of online communication these days. Communication in text form makes up a large part of it. The machine processing of these texts still poses a challenge in many cases.[1] This is the case with many languages, even with languages that have a large number of speakers who have had good resources and tools for decades. The number of resources and tools for the Albanian language is far from satisfactory. First and foremost, there is a lack of language resources. Even basic resources are missing, e.g. a free full-form lexicon. Most of them could be developed by linguists of the traditional schools. In this sense, the Albanian language is still considered an under-resourced language.
A large number of social media users use non-standard (lexical) variants of words instead of using the standard Albanian language. This often makes the text difficult to process, even to read and understand for the humans themselves.
We focus here on the use of non-standard lexical variants. Since there are many difficulties involved and they cannot be solved quickly, we are planning a somewhat longer-term project and are presenting an outline of the preliminary work here.

2. The lexical variation of Albanian

The Albanian language is mainly spoken in two varieties (dialect groups), in Gheg, spoken in the north, and Tosk, spoken in the south of the river Shkumbin. The dialects differ mainly phonetically, but are mutually intelligible, i.e. the respective speakers understand each other without difficulty. But that is only one aspect of lexical variation. Since Albanian is spoken in several countries and is taught under different school systems, it has developed several variations. The cultural and economic exchange of the respective countries where Albanian is spoken with neighboring countries or other countries has also contributed to lexical variation, e.g. in loan words (from Greek, Italian, South Slavic languages, as well as English, French and German) as well as in technical and scientific terminology. For example, a tool in Albania may have a different name than one in Kosovo – even if this has been standardized (in many cases) in the dictionary. In addition, it is well known that the social background and education of the language users contribute to lexical diversity. There are other factors that can play a role, but these cannot be dealt with here.

3. Computer-mediated communication

Computer-mediated communication (CMC), online communication, or communication at a distance, refers to *instant messaging*, *e-mailing*, *chatting*, *online forums*, *social networks* and similar services, or *social media* for short. It is easier and more common for users from different cities, regions, dialects and countries to come together and meet online than to have to meet in person (i.e. not online).

3.1. The use of the Albanian language in CMC

On the basis of empirical data, i.e. corpus data, and the word lists, frequency lists, n-grams, collocations, etc. extracted from them, we have identified various phenomena and analyzed them and developed machine language processing methods to deal with them. Some frequent ones are, above all, *dialect* or *idiosyncratic variants* of words, *abbreviations*, *contractions*, *emoticons*, and *creative spellings*. In the Albanian language used in the social media such word forms, e.g. the abbreviation *flm* (short form for *ju/të/… falemnderit*, engl. *thank you*), e.g. the contraction *ti* (engl. *you*) instead of *t'i* (i.e. subjunctive'accusative or dative object clitic), and e.g. the creative spelling *(e) ver8* instead of *(e) vertetë* (engl. (*it is*) *true*), i.e. *8* stands for *tetë* (engl. *eight*), can no longer be omitted in the communication texts.
An important feature is that they can be both *synchronous* and *asynchronous*. The synchronous communication is di-

[1] See especially here, among others, the series of CMC conferences, i.e. (Cotgrove et al., 2023) and the previous ones, which deal with different areas and topics, the EmiriST shared task, i.e. (Beißwenger et al., 2016) for linguistic annotation, as well as the *Journal of Computer-Mediated Communication (JCMC)*.

rect, fast and it often causes poorer text quality, e.g. more spelling mistakes. Some participants/users (i.e. 2nd or 3rd generation migrants) do not master the standard of the Albanian language and thus write in a deviant (sometimes, personal) version, which results in *spelling mistakes*. In addition, the speakers of the Albanian language, especially those of the 2nd and 3rd generation of migrated Albanians, very strongly *mix the code* with and use *words* and *idioms of the respective country language*, where they live, which results in *code mixing*. An easy, not difficult example for *dialect words / regionalisms* is *[. . .] sdi ca ke shkru [. . .]* instead of *s'di çka ke shkruar*, engl. (*I do not know what you have written*). Also, who writes to whom (all possible relationships can occur) determines the use of language, such as the choice of words, the style, etc. Often, CMC is seen as a very close version to spoken language. This is important for processing the syntax, i.e. for parsing.

The Albanian alphabet is based on the Latin alphabet with the addition of the letters *ë*, *ç* (with diacritics), and ten digraphs *dh*, *gj*, *ll*, *nj*, *rr*, *sh*, *th*, *xh*, and *zh*. The Albanian Language codes are *sq* (ISO 639-1), *alb* (*B*) (ISO 639-2), *sqi* (*T*) (ISO 639-3). The Albanian alphabet is covered by the ISO-8859-1 / Latin-1 character set (West European languages), and other code pages, e.g. by the ISO-8859-3 / Latin-3 character set (Southeast European languages) – and consequently also by Unicode.

The two characters *ë*, *ç* in particular cause problems, as they are often used in the version without diacritical marks. This causes ambiguities and makes language processing (more) difficult. This is because many users use different keyboards and input systems that do not offer the two letters with diacritics directly.

3.2. The collected linguistic data

We have been observing this type of communication ourselves for years and have been collecting texts for years. We have built up three relatively large corpora of Twitter data, i.e. one of them is standard language 9.2 million words, two of them are not-standard language, approx. 3.4 million words and approx. 0.8 million words.[2] We also have large amounts of data in Albanian from Reddit that are currently being processed.

4. The data processing and the NLP tools

We have the following working pipeline for processing the data: (1) *correct encoding* of the source texts/data, (2) *normalization* of the data, (3) *lemmatization*, and (4) the *POS-tagging*.[3] Normalizing the texts is the biggest challenge. We try to normalize the high-frequency cases first. The idiosyncratic variants in particular are very difficult to identify and therefore to normalize. We try to do this manually to create or improve the gold standard, as well as using various training models. The lemmatization (of standard lexical variants) is also not easy, since the Albanian language has a rich morphology and thus has a strong inflection, especially compounds, are difficult, but it is somewhat easier than the normalization. We normalize the non-standard variants first and consequently we lemmatize them. Our goal is to annotate the data according to the suggestions of EmpiriST Corpus 2.0, cf. (Proisl et al., 2020). We use the tagset from (Kabashi and Proisl, 2018), which also offers a mapping to UD tagset and Google tagset. We benefit from the gold standard and models created from it to tag the data. Based on this, we created new models with the collected CMC data. For gold standard creation, which is constantly being improved, two annotators are currently annotating. For variants, annotators need more linguistic knowledge than for standard variants. As this is a work in progress, we do not include preliminary results such as inter-annotator-agreement here. Based on this work and the experience gained, we want to adapt or extend the existing tools for language variants and create (language variant) resources[4] that make processing of CMC data easier and better.

We partly use *The IMS Open Corpus Workbench*[5] (*CWB*) tools for processing the data and CQPweb, cf. (Hardie, 2012) as a web platform for using corpora.

5. References

Beißwenger, M., Bartsch, S., Evert, S., and Würzner, K.-M. (2016). EmpiriST 2015: A shared task on the automatic linguistic annotation of computer-mediated communication and web corpora. In Paul Cook, et al., editors, *Proceedings of the 10th Web as Corpus Workshop (WAC-X) and the EmpiriST Shared Task*, pages 44–56, Berlin. Association for Computational Linguistics (ACL).

Louis Cotgrove, et al., editors. (2023). *Proceedings of the 10th International Conference on CMC and Social Media Corpora for the Humanities 2023 (CMC-2023)*, Mannheim, Germany. Leibniz-Institut für Deutsche Sprache (IDS).

Hardie, A. (2012). CQPweb - Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3):380–409.

Kabashi, B. and Proisl, T. (2018). Albanian part-of-speech tagging: Gold standard and evaluation. In Nicoletta Calzolari, et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Kabashi, B. (2018). A lexicon of Albanian for natural language processing. *Lexicographica*, 34(1):239–248.

Proisl, T., Dykes, N., Heinrich, P., Kabashi, B., Blombach, A., and Evert, S. (2020). EmpiriST corpus 2.0: Adding manual normalization, lemmatization and semantic tagging to a German web and CMC corpus. In Nicoletta Calzolari, et al., editors, *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020*, pages 6142–6148, Marseille, France. European Language Resources Association (ELRA).

[2] As far as we know, no other corpora of this type have been created for Albanian so far.

[3] See (Kabashi and Proisl, 2018) for an overview of the annotation tools as well as a tags set for the Albanian standard language. See (Kabashi, 2018) for a lexical resource.

[4] For example an extensive, well-covering list that maps various non-standardized lexical variants to standardized lexical variants.

[5] https://cwb.sourceforge.io.

Texting in Time: Approaching Processualities in Everyday Mobile Messaging Interaction

Jasmin Lallo
University of Bern
E-mail: jasmin.lallo@unibe.ch

Abstract

With the digitalisation of societies worldwide, smartphone-mediated communication has become integral to our daily routines and a subject of extensive research interest. However, our understanding of text production processes involved in everyday smartphone-mediated, text-based communication is limited due to challenges in accessing relevant data. This doctoral project, part of the Swiss National Science Foundation (SNSF) project Texting in Time, aims to provide novel perspectives on the processuality of everyday smartphone interactions.

The data consist of authentic screen recordings from participants in German-speaking Switzerland and Germany, documenting their smartphone communication on various social media and mobile messaging platforms over 14 consecutive days and participating in an ethnographic interview after completing the recording period. This methodologically innovative approach allows for the examination of smartphone-mediated interaction in terms of linguistic practices, the processual organisation of interactions, and patterns of platform choice. To investigate the interplay between these facets, a three-dimensional model of analysis is applied, comprising: (1) the writing process, (2) interactional practices, and (3) the patterned temporal organisation of participation. Drawing on Perrin's (2019) contributions within the framework of text production research, the first dimension encompasses the writing processes involving the smartphone keyboard interface. Building on Beißwenger's (2005) concept of *interaction management*, the second dimension focuses on how participants organise their interactions. Tagg and Lyons' (2021) concept of *device attention* frames the third dimension, which considers the timing and sequencing of participation within the rhythm of interactions shaped by social conventions and individual daily routines.

To analyse and effectively represent emerging patterns, a notation system and visualisation scheme are being developed based on existing transcription systems (Perrin, 2003; Kollberg & Severinson Eklundh, 2002). As part of the broader SNSF initiative, the project will culminate in an online database of mobile messaging interactions in video format, enabling future research across various disciplines.

Keywords: Building CMC corpora: from data collection to publication, Sociolinguistic studies of CMC, Multimodal (incl. visual) aspects of CMC

Kollberg, P., & Severinson Eklundh, K. (2002). Studying writers' revising patterns with S-notation analysis. In G. Rijlaarsdam, S. Ransdell, & M. Barbier (Eds.), *Contemporary tools and techniques for studying writing*, pp. 89--104.

Perrin, D. (2019). Irgendwie bin ich immer am Schreiben: vom Sinn transdisziplinärer Analysen der Textproduktion im Medienwandel. *Journal für Medienlinguistik*, 2(1), pp. 14--47.

— (2003). Progression analysis (PA): investigating writing strategies at the workplace. *Journal of Pragmatics*, 35(6), pp. 907--921.

Tagg, C., & Lyons, A. (2021). Polymedia repertoires of networked individuals. A day-in-the-life approach. *Pragmatics and Society*, 12(5), pp. 725--755.

Testing the behaviours of two laughter markers in a sample Twitter corpus: the "xD" emoticon and the "face with tears of joy" emoji

Sandra Marion
University of Caen Normandy, CRISCO
E-mail: sandra.marion@unicaen.fr

Abstract

Laughter, and the expression of laughter, while being frequent in daily conversations, has been fairly understudied from a linguistic and corpus standpoint, despite existing literature in the health and psychological fields. In computer-mediated communication (CMC) studies, research has been mostly carried out on the markers of laughter, such as *haha* or *lol* (Petitjean & Morel 2017, Schneebeli 2019), or on pictograms (Sampietro 2021). Following the existing research, this study aims to examine the functions and contextual usage of laughter markers in CMC, focusing specifically on the comparison and competition between the emoticon "xD" and its variants, and the emoji "😂". We aim to examine whether these two pictograms simply compete for the same linguistics functions (Pavalanathan & Eisenstein 2016), or if they have specific usages and meanings. We employ a usage-based approach, firstly through a distributional analysis identifying significant collexemes and patterns of usage, and then combining conversation analysis (Glenn 2003) and pragmatic analysis (Yus 2014) to compare the behaviour of these two laughter markers in online interactions. For this study, a preliminary corpus of English Twitter conversations was compiled in 2022 using *Social Bearing*. This provided us with 32 conversations (192 turns) which were used as a basis for a quantitative overview and a qualitative analysis. Preliminary results suggest that these two laughter markers tend to mirror the functions and placement of laughter in face-to-face interactions. The predominant function observed in our Twitter corpus is expressive, indicating reactions to humorous content. The results also suggest that "xD" is used in more face-preserving contexts than "😂". While this study provides insights on the behaviour and competition of two laughter markers on Twitter, further research, particularly employing multimodal approaches, is necessary to explore additional laughter markers and their combination with other linguistic material such as lexical selection or discourse type (see Wagener 2019, Smith 2022).

Keywords: laughter markers, computer-mediated communication, pictograms, Twitter corpus

References

Glenn, P. (2003). *Laughter in Interaction (Studies in Interactional Sociolinguistics)*. Cambridge: Cambridge University Press.

Pavalanathan, U., Eisenstein, J. (2016). More emojis, less :) The competition for paralinguistic function in microblog writing. *First Monday*. (doi:10.5210/fm.v21i11.6879)

Petitjean, C., Morel, E. (2017). "Hahaha": Laughter as a resource to manage WhatsApp conversations. *Journal of Pragmatics* 110. pp. 1--19. (doi:10.1016/j.pragma.2017.01.001)

Sampietro, A. (2021). The Use of the "Face with Tears of Joy" Emoji on WhatsApp: A Conversation-Analytical Approach. US: ICWSM. (doi:10.36190/2021.03)

Schneebeli, C. (2019). The meaning of LOL: patterns of LOL deployment in YouTube comments. *ADDA 2 – Approaches to Digital Discourse Analysis*, Turku, Finland.

Smith, C. A. (2022). *La combinatoire motivationnelle dans le lexique de l'anglais: Approche ascendante empirique des liens lexicaux en usage. Vol. Monographie inédite* [Habilitation, Université Lyon 3 J. Moulin].

Wagener, A. (2019). *Systémique des interactions : communication, conversations et relations humaines* (Questions contemporaines). Paris: l'Harmattan.

Yus, F. (2014). Not all emoticons are created equal. *Linguagem em (Dis)curso* 14. pp. 511--529. (doi:10.1590/1982-4017-140304-0414)

The 3DSeTwitch corpus – A three-dimensional corpus annotated for sexist phenomena

Ariane ROBERT
Université de Lille & STL CNRS UMR 8163, Lille, France
Paola PIETRANDREA
Université de Lille & STL CNRS UMR 8163, Lille, France, Institut Universitaire de France
E-mail: ariane.robert@univ-lille.fr, paola.pietrandrea@univ-lille.fr

Abstract

In this paper we present our work on the creation of the 3DSeTwitch corpus, a multimodal corpus aligning the representation of chats, audios and videos from Twitch, annotated for hate speech phenomena. Twitch is a platform for sharing live multimedia streaming experiences. It brings together internet users who interact live with each other – textually via chat, visually via video and verbally via audio. This Platform addresses cyber violence and sexist hate in particular. The creation of the corpus follows these different stages: (i) the data collection; (ii) the automatic extraction of data and metadata; (iii) the manual identification of samples associated with sexist language; (iv) the representation of the corpus with the CMC-core scheme; (v) the original annotation scheme of explicit and implicit sexist discourse; (vi) the inter-annotator agreement, carried out using a perspectivist approach.

Keywords: Twitch, sexism, CMC-core, three-dimensional corpus, gaming

1. Introduction

Twitch is a platform for sharing live multimedia streaming experiences that anticipates the new use of the internet as virtual worlds and immersive spaces.

Twitch allows users to use three different semiotic entities at the same time interval in a specific way. Streamers can broadcast live video from both their camera and their screen. Viewers can interact both with the streamer and with each other via a chat, *i.e.* an instant messaging window that can be used by both viewers and streamers. This creates a unique communication environment that is often referred to as "interactive television" (Recktenwald, 2017).

Twitch has faced significant issues related to cyber violence and sexist hate, in part due to Twitch's association with the gaming world, which is known for misogyny and sexism. According to a May 2023 report by the *NYU Stern Center for Business and Human Rights*, 47% of women and 37% of LGBTQIA+ people have been harassed on gaming platforms because of their gender. Influential streamers like Maghla have spoken out against this harassment and inspired others to do the same.

The NYU report also highlights the prevalence of extremist and hateful ideologies on platforms like Twitch and Discord, which foster autonomous, private communities due to their configuration. This makes the task of moderation difficult compared to traditional social networks such as X, Instagram or Facebook. Furthermore, users are unaware of the perlocutionary function that communication on the platform fulfils: while they think they are using these platforms to play games, they are taking a political message home with them.

By combining frameworks such as language and gender studies, critical feminist discourse analysis, corpus linguistics and CMC studies, we build the 3DSeTwitch corpus using a corpus-driven approach that addresses theoretical, methodological and applied questions, with the aim of identifying and classifying linguistic patterns of sexist discourse on Twitch.

2. The 3DSeTwitch corpus

2.1 Data collection

We have selected 46 videos and chats from 10 different channels of the most popular French male streamers: AntoineDaniel, Etoiles, Inoxtag, JLAmaru, Jolavanille, Kamet0, Mistermv, Ponce, RebeuDeter, and Tonton, and 50 videos and chats from 10 different channels of the most popular French female streamers: AVAMind, BagheraJones, DamDamLive, Deujna, Gom4rt, JeelTV, juliabayonetta, LittleBigWhale, Maghla, and Ultia. To create a popularity ranking and identify the most popular streamers, we used Twitch's statistical analysis programmes: Sullygnome [1] and Twitch Stat's [2] one. As female streamers only appear from the fiftieth or sixtieth place, we used exclusive rankings for women. We selected streams with at least 10,000 views, downloading no more than 5 recent streams per channel. We collected a total of 406 hours and 11 minutes collected from January 2022 to April 2022.

2.2 Automatic extraction

To extract the data automatically, we used the free opensource software TwitchDownloader, which is available on GitHub. It extracts videos in .mp4 format and chats in .txt and .json (.csv) formats. The .txt format provides the data (messages) and three metadata fields (the date and time of the message and the user name of the user). The .json or .csv format provides metadata for the channel and the video as well as metadata for each message and each viewer.

[1] https://sullygnome.com/
[2] https://gamingcampus.fr/tomorrow-lab/gaming-industry/gaming-campus-sassocie-a-twitch-stats.html

2.3 Manual identification of sexist language

We have manually identified content that refers to sexist language. Based on usage-based linguistics, this step is crucial for initial analyses and the development of an annotation scheme for automatic extraction. Our onomasiological approach starts from the functions (sexism) to find the forms that express them and does not rely on already established forms, as the phenomenon remains poorly described. The manual identification was based on chat readings, isolating as many sexist fragments as possible. We are aware that extraction at this stage is subjective, we will address this point further below.

2.4 Corpus representation

Twitch comprises three semiotic units simultaneously, which raises the question of how these layers can be represented and aligned.

We are adapting the CMC-core annotation scheme (Beisswenger and Luengen, 2020) for Twitch data. This schema structures linguistic data from digital communication in the TEI format (Text Encoding Initiative, 2019), an interoperable, openly accessible and well-documented standard. It can handle complex multimodal data with verbal (oral and written) and non-verbal elements. CMC-core has made four adaptations to TEI:

- Introduction of a new module called *cmc* for new classes and the *post* element;
- Addition of the new *post* element;
- Definition of the *model.CMC* class, which contains the post element as a member, aligning with *model.common*;
- Introduction of the new attribute class *att.CMC*, which defines the attribute *generatedBy*, with the *post* element and its children as members.

We are adapting the scheme in three respects. The tags <u>, <post>, <kinesic>, and <incident> are used to align verbal and non-verbal data. We want to annotate images and distinguish between the streamer's screen view and the camera view. We want to define different roles to distinguish between streamers, viewers, moderators and avatars. We also propose a classification of stream categories.

Figure 1: The annotation scheme

2.5 Annotation of sexist discourse

Based on various studies on recognising sexist content (Parikh *et al.*, 2019; West & Zimmerman, 1987; Culpeper, 2009; Karaian, 2014; Dragotto *et al.*, 2020; Zeinert *et al.*, 2021; Mojdehi, 2018; Minnema *et al.*, 2022; Bianchi, 2021; Campbell, 2015), we have created an annotation scheme of sexist language that can be identified at the semantic, syntactic, discursive and pragmatic levels (see Figure 1). While there are many classifications for direct sexist discourse in natural language processing, our aim is also to formalise indirect sexist discourse as described by Mills (2008).

2.6 A perspectivist approach

Our corpus includes clear instances of direct sexism, such as insults whose sexist nature is undeniable: *niquez vos mères* "fuck your mothers", *fils de pute* "son of bitch", *salope* "slut". It also contains examples where determining their sexist nature is more challenging:

(1) *Maghla c'est une douceur sur Internet, quelle femme*
Maghla is a sweetheart on the internet, what a woman.

Since sexism lacks a consensual definition and is subjectively perceived across various social and cultural contexts, we aim to evaluate the judgment regarding the sexist nature of our examples through an experimental procedure. By incorporating diverse viewpoints on statements deemed sexist or not, we can better capture the complexity of sexism, which a single researcher might overlook due to personal biases. This approach aligns with recent linguistic studies on perspectivism (Cabitza *et al.*, 2023), highlighting the importance of diverse perspectives to enhance objectivity and define the studied object, rather than relying on consensus.

2.7 Publication of the corpus and ethical considerations

We will use a representation format based on the CMC-core to ensure that the 3DSeTwitch corpus is interoperable and accessible to the scientific community.

To protect the data, we will anonymise the data during annotation using the CMC Core schema. Although anonymising multimodal data (oral and visual) is challenging, the Twitch setup only allows streamers to use video (image and sound). As we have selected popular channels and public figures, we will not anonymise the audiovisual data. The text data will be anonymised in accordance with the TEI guidelines: Each user will be anonymised with an identifier, keeping only the information important for our analysis (gender, if provided, and their role in the live event).

We will submit a file to the data protection department of the University of Lille to ensure the protection of the data.

3. Conclusion

We will deliver by the end of 2025 a multidimensional corpus of French video, audio and text excerpts from Twitch annotated for direct and indirect sexist language and represented by an extension of the CMC-Core scheme, whose annotation will be evaluated using a perspective approach.

We hope that this work will help to define standards for online moderation of sexist content and improve the legal definition of sexist discourse at the European level.

4. Copyrights

5. References

Beißwenger, M., & Lüngen, H. (2020). CMC-core: a schema for the representation of CMC corpora in TEI. *Corpus*, (20).

Bianchi, C. (2021). *Hate speech: Il lato oscuro del linguaggio,* Gius.Laterza & Figli Spa.

Cabitza, F., Campagner, A. & Basile, V. (2023). Toward a Perspectivist Turn in Ground Truthing for Predictive Computing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(6), 6860-6868. DOI : 10.1609/aaai.v37i6.25840

Campbell, E. (2015). Petite enquête lexicographique : au royaume des insultes, la femme est-elle l'égale de l'homme ? *French Studies Bulletin*, *27*(98), 13-16.

Culpeper, J. (2009). *Impoliteness: Using and Understanding the Language of Offence*, British Economic and Social Research Council Project Website, http://www.lancs.ac.uk/fass/projects/impoliteness

Dragotto, F., Giomi, E. & Melchiorre, S. M. (2020). Putting women back in their place. Reflections on slut-shaming, the case Asia Argento and Twitter in Italy, *International Review of Sociology*, 30(1), 46-70, DOI: 10.1080/03906701.2020.1724366.

Karaian, L. (2014). Policing "sexting": Responsibilization, respectability and sexual subjectivity in child protection/crime prevention responses to teenagers' digital sexual expression. *Theoretical Criminology*, 18(3), 282–299.

Mills, S. (2008). *Language and Sexism*. Cambridge, U.K., Cambridge University Press. https://doi.org/10.1017/CBO9780511755033

Minnema, G., Gemelli, S., Zanchi, C., Caselli, T. & Nissim, M. (2022). *Dead or Murdered? Predicting Responsibility Perception in Femicide News Reports.* DOI : 10.48550/arXiv.2209.12030

Mojdehi, A. L. (2018). Stéréotypes de genre et sexisme : principaux registres d'insultes dans les espaces publics. *Cahiers du genre*, 65(2), 169-191.

Pardo, L. (2019). Twitch Dowloader (1.54.2) [software]. Github. https://github.com/lay295/TwitchDownloader

Parikh, P., Abburi, H., Badjatiya, P., Krishnan, R., Chhaya, N., Gupta, M. & Varma, V. (2019). Multilabel categorization of accounts of sexism using a neural framework. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP),* 1642–1652, Hong Kong, China. Association for Computational Linguistics.

Recktenwald, D. (2017). Toward a transcription and analysis of live streaming on Twitch. *Journal of Pragmatics*, *115*, 68-81.

Rosenblat, M. O. (maiy2023). *Gaming The System: How Extremists Exploit Gaming Sites And What Can Be Done To Counter Them.* NYU Stern Center for Business & Human Rights. https://bhr.stern.nyu.edu/publication/gaming-the-system-how-extremists-exploit-gaming-sites-and-what-can-be-done-to-counter-them/

Text Encoding Initiative (2019). « Text Encoding Initiative Consortium » [website] [https://teic.org/].

West, C. & Zimmerman, D. (1987). Doing Gender. *Gender and Society*, 1(2), 125-151.

Zeinert, P., Inie, N., & Derczynski, L. (2021). Annotating online misogyny. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, Association for Computa- tional Linguistics, Online*.

Communication Dynamics in 'No Vax' Groups During Deradicalization Phases

Andrea Russo
Sorbonne University & CNRS
Andrea.russophd@gmail.com

Abstract

This paper tries to enlighten the communicative variations in the world of No-Vax on telegram during a GPT interaction to deradicalise the group. Using computational methods, different approaches were developed to visualise the communicative dynamics before and after a deradicalisation phase. The deradicalisation process was carried out by various bots under the control of an AI. The aim of the paper is to observe the temporal variations in communication before and after the AI's deradicalising intervention.
Keywords: Sociology, Computational Methods, Word-Embedding, Telegram

1. Introduction

Radicalization, defined as adopting extreme ideologies condoning violence for political or ideological purposes (Serafim, 2005), is increasingly prevalent within digital networks. Platforms like Telegram serve as central nodes for spreading inappropriate and illegal content, leveraging the ease of sharing and perceived anonymity these platforms offer (Semenzin and Bainotti, 2020). Algorithms further exacerbate this by facilitating rapid dissemination and the creation of deceptive content such as deepfakes. While digital communities are well-studied, there is limited focus on communication dynamics during deradicalization. This paper addresses this gap through an action-research study that uses computational techniques to analyze communication changes during deradicalization. Radicalization is a phased process where individuals or groups embrace ideologies condoning violence for political or ideological goals (Serafim, 2005). It can occur at individual, social, and governmental levels, with no single pathway but various forms influenced by multiple triggers (European Commission, 2024a; European Commission, 2024b). The Italian No Vax group studied here is highly active, with significant radicalization (Russo, 2023). This group, involved in damaging public property and threatening individuals, includes radicals promoting violent actions against vaccinations. The digital era has expanded opportunities for disseminating extreme content, with online platforms used for recruitment and psychological manipulation (Blood, 2024; Area, 2024). Despite being smaller, the Italian anti-vax community is highly cohesive and isolated from pro-vax influencers, making pro-vax campaigns less effective (Francia et al., 2019). Anti-vax communication strategies are effective due to their simplified, emotional narratives, blending safety concerns, conspiracy theories, and alternative medicine (Herasimenka et al., 2023; Bianchi and Tafuri, 2023). These groups use digital platforms to enhance organizational capabilities, incorporating traditional propaganda and new models like "digital celebrity" (Herasimenka et al., 2023).

Methodologies such as word embedding and sentiment analysis (Russo, 2023) or semantic networks based on word frequencies (Russo et al., 2024) are valuable for observing communication changes over time (Jeon, 2023). The deradicalization process, where individuals renounce extreme ideologies, is complex and influenced by group identity and inter-group dynamics (Doosje et al., 2016). Understanding communication differences can reveal social dynamics and resistance to deradicalization. This research aims to explore these aspects, filling a gap in the literature on deradicalization in online radical groups.

2. Data & Methodology

The group *'Put down the covid-mask'* has about 450 people (2022/11/28) with an average of about 70 people online daily for a total of about 1300 interactions (Russo, 2023). Data were collected pre-interaction with GPT, and post-interaction to observe group structural differences (Russo, 2023). Analyses of interactions and communications were carried out using various tools and methods, such as sentiment analysis with the Italian VADER, word-embedding and word networks (in this case to highlight various related concepts) (Russo, 2023; Russo et al., 2024). I used three different methods for analysing the text to get a better picture of the situation on different points of view, as sentiment analysis gives a value on content, Bigram on word relations, and embedding on spatiality and difference between words (Russo et al., 2024).

2.1. VADER Sentiment Analysis

VADER can determine if text expresses positive, negative, or neutral sentiment, along with the intensity of each sentiment (Russo, 2023). VADER was originally developed with an English lexicon, but there have been efforts to adapt it to other languages, including Italian (Russo, 2024). An Italian lexicon has been curated by myself specifically for sentiment analysis using VADER, allowing it to analyze sentiment in Italian text more accurately with VADER (Russo, 2024).

2.2. Bigram and Words-networks

After sentiment analysis, another effective method involves analyzing word frequency within sentences and establishing connections between them (Russo et al., 2024). This approach identifies the most frequently used words in each sentence and examines their relationships, mapping out connections between words within the same message. This analysis reveals contextual relationships between specific concepts or topics and other words (Russo et al., 2024).

The final result shows how certain concepts or topics are

Table 1: Average text sentiment of the Telegram group before and after dynamics

Timing	Negative	Neutral	Positive	Compound
Initial condition	0,125332436	0,724032301	0,149318977	0,481011844
After dynamics	0,078551515	0,938115152	0,104549495	0,067783434
Variation	-0,046780921	0,214082851	-0,044769482	-0,41322841

linked to other words, creating the context of the topic and improving comprehension (Russo et al., 2024). To link words within each sentence, the Bigram tool identifies the most frequently used words (as nodes) and connects them (with edges) to the second most frequently used words within the same sentence. The result is a network of words (Russo et al., 2024).

2.3. Word-Embedding

A word embedding is a representation of words in a continuous vector space where words with similar meanings are mapped to similar vectors. The mapping is learned from data, and in our case I used neural network models (Word2Vec) (Russo, 2023; Jeon, 2023). Word embeddings capture semantic relationships between words, enabling algorithms to process language more effectively by understanding context and meaning (Jeon, 2023). In the end of the embedding process, we would obtain a two-dimensional concept map, where each word would be placed within the map based on its position on the proximity or distance of similarity with other words in the corpus (Jeon, 2023).

In the end, the deradicalisation phase was successful in changing the structure of the networks, creating a polarisation that caused the administrator to close the Telegram group (Russo, 2023). The dynamics of deradicalisation caused uproar in the Telegram group, moving from what is purely hate speech to a more neutral language. Table 1 shows the sentiment differences obtained before and after the transition phase (Russo, 2023).

3. Result

The content analysis shows that before deradicalization, the group's sentiment was highly polarized with notably high negative and positive values, often achieved through ironic phrases (Russo, 2023). There was a substantial presence of comments endorsing weapons, revolt, and conspiracy theories involving figures like George Soros and Klaus Schwab, while post-deradicalization, neutrality levels improved significantly, and both positive and negative sentiments decreased, indicating reduced ironic and hateful content (Russo, 2023). The network results show significant structural and content differences. Figure 1a depicts a complex network with loosely connected topics sharing common words, while Figure 1b shows a linear and closely linked series of topics (Russo, 2023; Russo et al., 2024). The first network is dominated by war discussions, whereas the second focuses on national politics and the economy. The AI-bots aimed to shift discussions from anti-vaccine and brainwashing topics to broader institutional and political subjects (Russo, 2023). Word-embedding results, similar to the word networks, show war as a main topic, with no-vax and USA/Ukraine-Russian conflict appearing prominently before deradicalization (Russo et al., 2024). Afterward, the focus shifted to economic inflation and national policy, confirming a subject change (Russo, 2023).

Area, R. (2024). Radicalisation in the digital era — rand.

Bianchi, F. P. and Tafuri, S. (2023). Spreading of misinformation on mass media and digital platforms regarding vaccines. a systematic scoping review on stakeholders, policymakers, and sentiments/behavior of italian consumers. *Human Vaccines & Immunotherapeutics*, 19(2):2259398.

Blood. (2024). Radicalisation and extremism — cambridgeshire and peterborough safeguarding partnership board.

Doosje, B., Moghaddam, F. M., Kruglanski, A. W., De Wolf, A., Mann, L., and Feddes, A. R. (2016). Terrorism, radicalization and de-radicalization. *Current Opinion in Psychology*, 11:79–84.

European Commission, E. (2024a). Prevention of radicalisation - european commission.

European Commission, E. (2024b). Prevention of radicalisation - european commission.

Francia, M., Gallinucci, E., and Golfarelli, M. (2019). Social bi to understand the debate on vaccines on the web and social media: unraveling the anti-, free, and pro-vax communities in italy. *Social Network Analysis and Mining*, 9:1–16.

Herasimenka, A., Au, Y., George, A., Joynes-Burgess, K., Knuutila, A., Bright, J., and Howard, P. N. (2023). The political economy of digital profiteering: communication resource mobilization by anti-vaccination actors. *Journal of Communication*, 73(2):126–137.

Jeon, S. (2023). Migrtwit corpora.(im) migration tweets of french politics. In *Proceedings of the 10th International Conference on CMC and Social Media Corpora for the Humanities*. University of Mannheim; Leibniz-Institut für Deutsche Sprache (IDS).

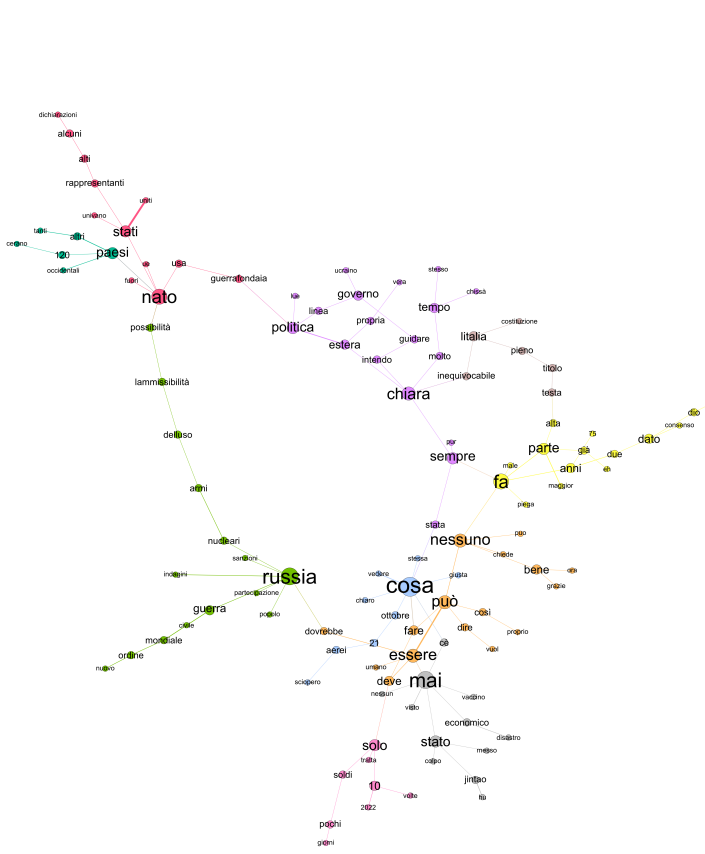
Russo, A., Miracula, V., and Picone, A. (2024). Topics evolution through multilayer networks; analysing 2m tweets from 2022 qatar fifa world cup. *arXiv preprint arXiv:2401.12228*.

Russo, A. (2023). Ai-influencer to mitigate radicalism and social threats on telegram.

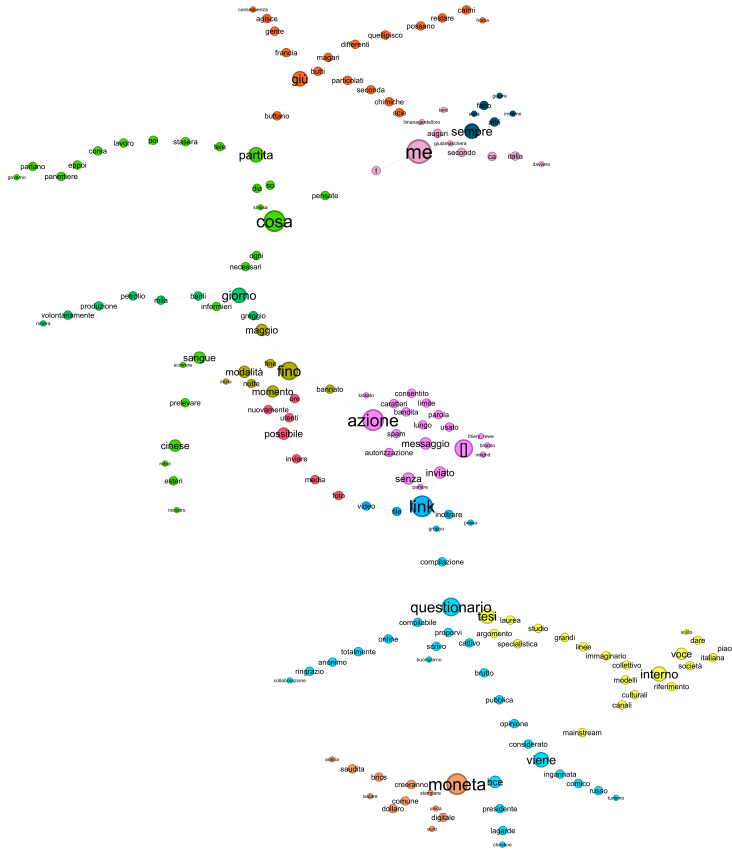
Russo. (2024). Github - andrearussoagid/vader-italian-sentiment.

Semenzin, S. and Bainotti, L. (2020). The use of telegram for non-consensual dissemination of intimate images: Gendered affordances and the construction of masculinities. *Social Media+ Society*, 6(4):2056305120984453.

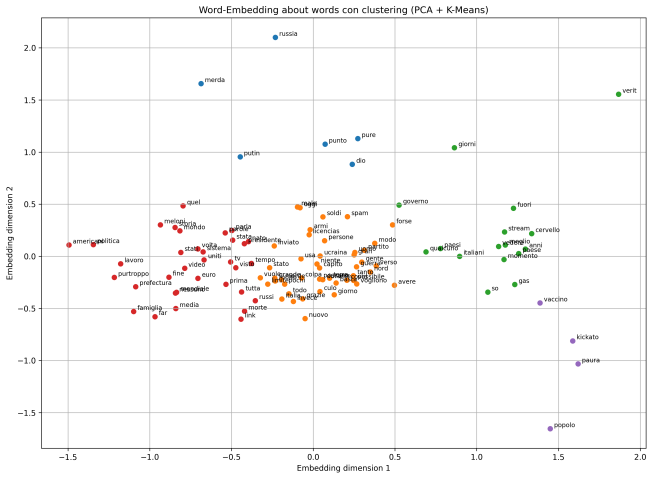
Serafim. (2005). Radicalization, defined as the adoption of extreme ideologies condoning violence for political or ideological purposes.



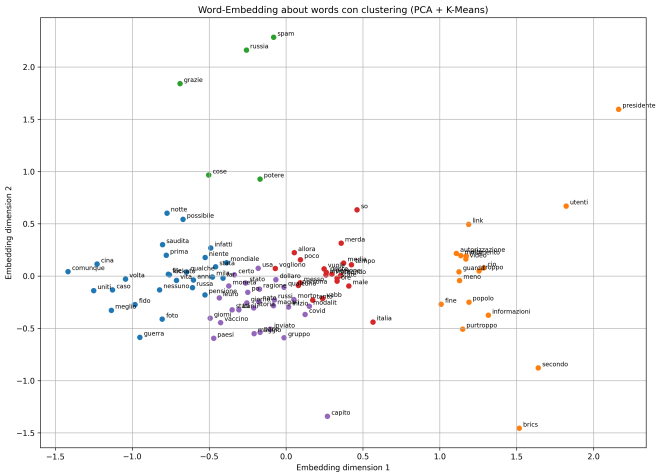
(a) Words-networks before deradicalization



(b) Words-networks after deradicalization



(c) Word-Embedding before deradicalization



(d) Word-Embedding after deradicalization

Figure 1: Comparison of Words-networks and Word-Embedding before and after deradicalization

Exploring discursive multidimensionality and multimodality on Twitter: Analyzing xenophobic representations targeting China during the COVID-19 pandemic

Cícero Soares da Silva
São Paulo Catholic University- PUC/SP
E-mail: pardonmester@gmail.com

Abstract

In this paper, we look at malicious representations of China on Twitter occurring in the context of the COVID-19 pandemic. The pandemic has heightened ongoing prejudice against China, its people and culture partly because the coronavirus originated in that country, and partly because the West has engaged in demoralizing campaigns against China for decades. In order to capture these detrimental representations, we scraped two corpora of ca. 100K tweets in Brazilian Portuguese containing ten highly xenophobic hashtags, namely #viruschines, #víruschines, #víruschinês, #viruschiunês, #pragachinesa, #pestechinesa, #pasteldeflango, #boicoteachina, #vachina, #ChainaVirus, and #wuhanvirus. These were used by right-wing followers in Brazil to discredit China and spread hatred. The multimodal method followed in this study consisted of an application of Lexical Multidimensional Analysis - LMDA - (Berber Sardinha & Fitzsimmons-Doolan, 2024) and Visual Multidimensional Analysis - VMDA - (Berber Sardinha et al., 2023). The LMDA used lexical units to detect traces of discourses across the texts, whereas the VMDA applied computer vision AI techniques to annotate the images posted along with the twitter messages. Two sets of dimensions were obtained: a verbal set:6 dimension, and a visual set: 5 dimensions denoting ideological positions aligned with anti-Chinese xenophobia during the COVID-19 pandemic.

Keywords: Xenophobia, Discourse analysis, Lexical Multi-Dimensional Analysis, Multimodal, Multi-Dimensional Analysis.

1. Introduction

This study seeks to describe the dimensions of discursive and ideological variation by analyzing the co-occurrence patterns of lexical discursive elements and multimodal (lexical and visual) content on Twitter, focusing on infodemic discourses related to xenophobia targeting China, Chinese people, and its culture during the COVID-19 pandemic in Brazil. Lexical MD Analysis (LMDA) is an extension of the original MD Analysis framework designed to identify the lexical dimensions of variation in a corpus. LMDA has been in development since the 2010s, primarily as an approach for the study of discourse, offering tools to analyze constructs like ideologies, representations, identities, and themes. The Corpora of Anti-Chinese Xenophobia (COXAC) were analyzed using multimodal multidimensional analysis (Berber Sardinha, 2022b), which comprise the analysis of the text and the visual components separately, with the goal of unveiling the major discourses emerging from the texts and the major visual characteristics present in the images, followed by the joint analysis of both modes.

2. Foundational Principles

Multi-dimensional (MD) Analysis, originally termed Multi-feature Multi-dimensional Analysis, was developed by Biber in the 1980s, initially to study variation between written and spoken text varieties and model register variation. In MD Analysis, register is typically understood 'as a cover term for any variety associated with particular situational contexts of purposes', which often correspond to 'named varieties within a culture, such as novels, letters, editorials, sermons, and debates', and which 'can be defined at any level of generality' (Biber, 1995, p. 1). The approach has since branched out to encompass further types of analyses, now covering the analysis of lexical features through Lexical MD Analysis, visual features via visual MD Analysis, and multimodality through Multimodal MD Analysis. Each of these MD types serves a unique research goal; however, they all adhere to the foundational assumptions of the original MD Analysis framework.

3. MD Analysis Methodology

We scraped two corpora of ca. 100K tweets in Brazilian Portuguese containing ten highly xenophobic hashtags, namely #viruschines, #víruschines, #víruschinês, #viruschiunês, #pragachinesa, #pestechinesa, #pasteldeflango, #boicoteachina, #vachina, #ChainaVirus, and #wuhanvirus to discredit China and spread hatred. The analysis was carried out using methodological extensions of multidimensional analysis—specifically, lexical, visual, and multimodal multidimensional analysis. Furthermore, this study includes information on tools for collecting and visually annotating a corpus of texts and images.

4. Results and Data Analysis

This section presents the findings related to the analysis of the six lexical identified factors and five visual identified factors.Two sets of dimensions were obtained, a verbal set, and a visual set. All dimensions were interpreted, labeled, discussed, and illustrated in the thesis. The working hypothesis was confirmed: each pole of the detected dimensions indicates a discourse (or a set of compatible discourses) denoting ideological positions aligned with anti-Chinese xenophobia during the COVID-19 pandemic.

4.1. Tables

Corpora used in this research are named the Anti-Chinese

Xenophobia Corpora - COXAC, as shown in Tables 1 and 2 below.

Year	Tweets	Lexical itens	Mean	Standard Deviation
2020	73.781	1.525.503	20.6	13.5
2021	7.080	1.549.46	21.8	13.9
2022	795	16.959	21.3	15.3
Total	81.656	1.697.408		

Table 1: Verbal *Corpus* Design

Year	Images	Labels
2020	11.478	11.477
2021	1.278	1.275

Table 2: Desenho do *Corpus* Visual

5. References

Berber Sardinha, T., & Fitzsimmons-Doolan, S. (2024). Lexical Multidimensional Analysis. Cambridge Univesrity Press.

Berber Sardinha, T. (2022b). *Patterns of lexis, patterns of text, patterns of images*. Talk presented at the Pathways to Textuality Meeting in Honor of Michael Hoey, Eastern Finland University.

Berber Sardinha, T., (2024b). Going multimodal in corpus linguistics: The case of social media. In. P. Crosswaithe (ed.), Corpora for Language Learning: Bridgiung the Research-Practice Divide. Abingdon: Routledge.

Biber, D. (1995). *Dimensions of Register Variation - A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.

Author Index