Register and Social Media

Isobelle Clarke
Lancaster University

Social media has become an integral part of modern-day life. Nearly 57 percent of the 7.7 billion people worldwide use it (4.48 billion) with the average social media user having 8.4 social media accounts and engaging with an average of 6.6 social media platforms (Dean 2021). Each social media platform varies in terms of the medium's characteristics and affordances (e.g., message format, privacy settings, synchronicity of message), and the situational factors (e.g., participants, goals, topics and norms). These register distinctions inevitably influence the language used on the platforms. Despite social media hosting a variety of interactional situations and therefore providing linguists with the opportunity to understand the social nature of language, social media have been underexplored from a register perspective.

One of the main reasons for this is because social media posts are typically short texts. To describe patterns of register variation on any scale, linguists often compare the normalised frequencies of features across texts. However, the normalised frequencies of features in short texts are not very meaningful. Given the brevity of social media texts, the raw counts of features in these texts are inflated in normalised rates to levels not seen in longer stretches of texts. For instance, a 5-word text with one adjective has adjectives occur 200 times per a thousand words. Yet even the most descriptive texts come nowhere near this rate. Likewise, short texts tend to have many features missing. Compare the sentences:

> (1) That man is very silly.
> (2) Yes, he is very silly.

While there is no doubt that both sentences are functionally and situationally very similar, they do not share two features. Sentence (1) does not have the interjection *yes* or the third person pronoun *he* and sentence (2) does not have the demonstrative determiner *that* or the noun *man*. These features would have a normalised frequency of 0 per a thousand words in the texts where they are absent, yet they will have a normalised frequency of 200 times per a thousand words in the text where they occur. These inflated disparities are not helpful in text comparisons. As a result, register studies of social media have been limited to longer texts above a particular length or several short posts concatenated to form longer text units that are more suitable for frequency-based analyses. This has limited the kinds of research questions to exploring register variation across platforms and not within platforms.

The development and introduction of short text approaches, such as by Clarke and Grieve (2017, 2019), Clarke (2020) and Berber Sardinha (2022, this issue), as well as studies considering the influence of text length on register in more detail (e.g., Liimatta 2020, this issue) are beginning to facilitate the large-scale analysis of social media from a register perspective. Yet there is still work to be done. This special issue of *Register Studies* brings together a collection of articles that are focused on the importance of register in social media. The articles explore the influence of a variety of register distinctions on different levels of language use, including a selection of the platforms themselves (Berber Sardinha; Biri; Marko, Reitbauer & Pickl; Scheffler, Kern & Seemann), subsidiary threads within a platform (i.e. subreddits) (Liimatta), group beliefs across platforms (Biri), differing perspectives and expectations within a single platform and genre (Veselovsky & Witzlack-Makarevich), the same communicative purposes across platforms (Scheffler et al.), authorial style across platforms (Marko et al.) and text types across platforms (Berber Sardinha).

Tony Berber Sardinha's article *A Text Typology of Social Media* introduces an initial text typology of social media posts. Using a short text version of multi-dimensional analysis (MDA) (Berber Sardinha 2022) on a corpus of English social media messages from Reddit, Twitter, Facebook, Telegram, Instagram, and YouTube, Berber Sardinha revealed three dimensions of linguistic variation. The dimension scores of each text were then used in a k-means cluster analysis to reveal two distinct text types. These text types are similarly found across other English texts, which indicates that social media most commonly draws on forms of expression outside of digital communication. His analysis finds that the different topics, groups, and users within each platform were better predictors of the variation than the platform alone.

Tatjana Scheffler, Lesley-Ann Kern and Hannah Seemann's article *The Medium Is Not the Message: Individual Level Register Variation in Blogs vs. Tweets* distinguishes register (here used to refer to communicative purpose) from medium and analyses the influence of registers (texts coded as narrative, informative and persuasive) and social media (blogs vs. tweets) on 44 parent bloggers' use of German modal and intensifying particles. Their study reveals that modal particles vary across media and across registers, whilst intensifiers vary only by register and not by the medium. Scheffler et al. therefore argue that language variation does not only depend on the medium, but also, and arguably more distinctively, on the register. They thus critique research which conflate register with medium and call for future research to acknowledge and include the different registers within linguistic analyses of social media texts.

In a different investigation of the same authors across registers (here defined as social media platform), Karoline Marko, Margit Reitbauer and George Pickl's paper *Same Person, Different Platform: Challenges And Implications for Forensic Authorship Analysis* explore the influence of register on the individual style of three authors across two social media platforms in order to generate hypotheses to be investigated on a much larger scale. Marko et al.analyse the three authors' social media posts for 52 linguistic features across different levels of language. The analysis reveals that for these three authors some features are stable across the platforms, whilst other features appear to be strongly influenced by the register. Consequently, Marko et al. call for more research on cross-platform forensic authorship analysis to investigate these patterns further.

Ylva Biri's article *Epistemic Stance in the Climate Change Debate: A Comparison of Proponents And Sceptics on Twitter And Reddit* analyses the influence of platform and climate change beliefs on register indexed through the use of epistemic stance. The study demonstrates that both the platform and group allegiance (proponent vs. sceptic of climate change) influence the ways in which certainty and veracity are evaluated. Specifically, climate change proponents discuss the likelihood of climate events, whereas climate change sceptics discuss the veracity of the claims. Reddit users use more hedging, whereas Twitter users categorically assert statements about climate science or claims based on opinion. Biri argues that this is indicative of the situational constraints and affordances of the platforms. Reddit allows for longer dialogical discussions, which may invite more interpersonal resources, whereas tweets are restricted in length to 280 characters, which may limit the number of interpersonal resources that can be included. Moreover, the tweets she analysed were less dialogical, exhibiting a more informational style, as described by Clarke (2022), due to their inclusion of hashtags.

Given that the length of a text is intrinsically linked with its linguistic aspects (e.g., the requirements and constraints of the situation and the communicative goals), Aatu Liimatta's paper *Do Registers Have Different Functions for Text Length? A Case Study of Reddit* considers the association of text length with different communicative functions. Although previous research (e.g., Clarke & Grieve 2017; Liimatta forthcoming) has revealed

a general trend of particular lengths of texts having particular communicative functions (e.g., longer texts are generally more informational and narrative, whilst short texts are more interactive and involved), Liimatta's analysis sought to explore how the functions associated with texts of a particular length vary between registers. For example, are shorter texts in some registers more narrative and are longer texts in other registers more involved? His analysis compares texts of different lengths within numerous subreddits. The analysis shows that the register category, here defined as the subreddit, can have a major influence on the functions which are associated with particular lengths of text. With short and long texts serving similar communicative functions in different subreddits, Liimatta demonstrates that the role of text length often depends on the register with some registers demanding longer texts for particular functions, whereas other registers demand shorter texts for the same functions. His analysis demonstrates that taking text length into consideration in register studies can be insightful.

Drawing on register analysis and narrative analysis (Labov & Waletzky 1967), Anna Veselovsky and Alena Witzlack-Makarevich's article *Linguistic Variation in Customer Reviews: One's Own vs. Another's Experience Narrative* examines customer reviews of assistive reading devices to explore intra-register variation between narratives of one's own and another's experience. The study shows that the distributions of particular linguistic features, such as pronoun use and stance markers, differentiate the two experience types. They suggest that the narrative ownership and the associated distinctions in expectations determine the use of particular linguistic features. Veselovsky and Witzlack-Makarevich thus call for more research to shed light on how narrative ownership affects other types of services to assess the extent to which their findings are universal to one's own and another's experience.

Despite all the articles in this issue taking different perspectives and using different approaches, they all demonstrate the importance of register on language use across social media. The articles provide new insights and directions for future linguistic research to further understand the social nature of language. We hope that these articles will encourage more research that foregrounds the importance of register when describing language on social media.

References

Berber Sardinha, T. (2022). Corpus linguistics and the study of social media: a case study using multi-dimensional analysis. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (2nd ed., pp. 656-674). New York: Routledge.

Clarke, I. (2020). *Linguistic variation across Twitter and Twitter trolling* [Unpublished doctoral dissertation]. University of Birmigham.

Clarke, I. (2022). A Multi-dimensional analysis of English tweets. *Language and Literature.* Advance online publication. doi:10.1177/09639470221090369

Clarke, I. & Grieve, J. (2017). Dimensions of abusive language on Twitter. In *Proceedings of the First Workshop on Abusive Language Online* (pp. 1-10). Vancouver: Association for Computational Linguistics.

Clarke, I. & Grieve, J. (2019). Stylistic variation on the Donald Trump Twitter account: A linguistic analysis of tweets posted between 2009 and 2018. *PLOS ONE, 14*(9), e0222062. doi:https://doi.org/10.1371/journal.pone.0222062

Dean, B. (2021, October 10). Social network usage & growth statistics: How many people use social media in 2022? Backlinko. Retrieved from https://backlinko.com/social-media-users

Labov, W., & Waletzky, J. (1967). Narrative analysis. In J. Helm (Ed.), *Essays on the verbal and visual arts* (pp. 12–44). Seattle/ London: University of Washington Press.

Liimatta, A. (2020). Using lengthwise scaling to compare feature frequencies across text lengths on Reddit. In S. Rüdiger & D. Dayter (Eds.), *Corpus approaches to social media* (pp. 111-130). Amsterdam/Philadelphia: John Benjamins.

Address for Correspondence

Isobelle Clarke
Department of Linguistics and English Language
County South
Lancaster University
Lancaster
United Kingdom
LA1 4YL

i.clarke@lancaster.ac.uk