

Linguistics of Social Media: An Introduction to the Special Issue

Andreea Simona Calude 

School of Psychology and Social Sciences, Division of Arts, Law, Psychology and Social Sciences, University of Waikato, Hamilton 3216, New Zealand; andreea.calude@waikato.ac.nz

It is hard to remember a world without social media (SM). The Internet Relay Chat (IRC) and bulletin boards of the 1980s and the popular early platforms of the 1990s, MySpace, AOL Messenger, and Facebook, gave rise to nothing short of an SM empire in just 50 years. Although we have been considering how language is used on the Internet for some time now, see [Crystal \(2001\)](#) for an initial example, studying the language used on SM is comparatively a newer field of inquiry.

One defining characteristic of social media is the dual role that users assume, being both consumers and producers of content ([Seargeant & Tagg, 2014](#); [Calude, 2023](#)), at least in theory if not always in practice. Even though lurkers are largely silent, they can nevertheless still influence the language used by their very presence as auditors ([Bell, 2001](#)) and as spectators, with the option of becoming contributors at any moment. This participatory culture ([Androutsopoulos & Tereick, 2015](#); [Jenkins, 2006](#)) adds to the possibility of a multitude of voices, synchronous and asynchronous commentary, responses, reactions, and the emergence of a multi-authored discourse. Nevertheless, it is worth echoing the caution stated by Jenkins that “not all participants are made equal” ([Jenkins, 2006](#), p. 3), with some voices remaining under-represented and silenced. Be that as it may, characteristics shift SM language beyond internet language, towards a more dynamic, collaborative and fast-paced channel of communication.

Current estimates suggest that roughly 65% of the total world’s population or around 5 billion people around the world are social media users (5.32 billion people, according to the Global Digital Overview produced by [Meltwater \(n.d.\)](#), and 5.24 billion people, according to [Statista, 2025](#)). With there being a platform for just about everything, from keeping in touch with friends and family (Facebook), to professional networking (LinkedIn), dating (Tinder), gaming (Discord), language learning (Duolingo), interest groups and image posting (Instagram) or video sharing (TikTok and YouTube), it is no wonder that language scholars have eventually found themselves on social media, documenting, describing, and analysing the language used in these virtual contexts. Added to that, the increased pressure to be “always on” ([Baron, 2008](#)) has made SM a bustling place of activity, including language activity.

Utilizing data from social media to deepen our understanding of language is profitable for at least two reasons. First, this communication channel constitutes a large, dynamic and constantly growing body of interactive data, which varies with respect to its accessibility across time and platform. Both its size and accessibility are notable advantages in a world where empirical, data-driven methods of language analysis dominate.

Social media provides language scholars with access to language-in-use on an unprecedented scale. As the articles included here illustrate, such vast data presents the opportunity to tap into language phenomena that have been previously difficult to scrutinize in sufficient detail, either because they are rare or because they occur in specific contexts that are not easily accessible to researchers. The relative ease of access is also



Received: 5 June 2025
Accepted: 9 June 2025
Published: 12 June 2025

Citation: Calude, A. S. (2025). Linguistics of Social Media: An Introduction to the Special Issue. *Languages*, 10(6), 140. <https://doi.org/10.3390/languages10060140>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

significant in allowing researchers to focus more of their time and efforts on questions and analyses than on data collection and transcription.

Secondly, although SM language is itself extremely varied, much content produced and digested online is posted spontaneously, off-the-cuff and with little planning and editing. It is language produced by everyday speakers and not just those whose jobs, positions or training grants them the possibility to do so (journalists, commentators, politicians¹). It is the language of commuters, of mum groups, of running enthusiasts and of Indigenous language activists, among many others. This kind of language is especially difficult to tap into and all the more valuable for its diversity of voices and styles. Spontaneous language can inform important debates about language change and provide a glimpse of how our minds handle language. It also shows usage as it happens rather than as we are editorially instructed or compelled to shape it (see similar arguments made in relation to spontaneous spoken language in [Miller & Weinert, 1998](#), and others).

Naturally, SM data does not come without its perils and limitations. As is evident from the methods of the articles included here and SM research elsewhere, extracting, cleaning, and sorting through social media data is not a trivial process. SM data is extremely noisy, and it is increasingly difficult to disentangle human texts from text produced by bots or other automated tools. Moreover, posts tend to provide little context and background for interpreting the message, leaving the researcher in the lurch about wider meanings and in-group uses. Moreover, the range of languages and dialects present on social media continue to fall short of global linguistic diversity, mimicking in digital spaces, the kinds of inequalities in resource distributions that we see offline ([Bender et al., 2021](#); [Deumert, 2015](#)).

Previous research scrutinizing language on social media has used discourse analytic tools to illuminate how SM users leverage language to construct and reflect their identity, how they present themselves online, and how they build virtual networks by utilizing affordances offered by various platforms, such as hashtags, hyperlinks, memes and @ mentions (see for example [Lee & Barton, 2013](#); [Newton et al., 2022](#); [Page, 2012](#); [Seargeant & Tagg, 2014](#); [Shifman, 2014](#); [Zappavigna, 2015](#), inter alia). Building on this body of work, the present collection of papers aims to provide a linguistically grounded lens of how communication unfolds on social media. At the same time, my intention was to seek work harvesting data from a varied sample of languages (including users of non-WEIRD languages) and a wide range of linguistic sub-fields, grammar, morphology, semantics and so on.

The resulting Special Issue contains analyses of Māori (an Indigenous language), French and Belgian French, features of African American Vernacular English and feminine-coded personae, German, and English. Analyses employ theoretical notions from grammar, pragmatics, sociolinguistics, and language contact. Short overviews of each of the five contributions are provided in the paragraphs that follow.

Trye et al. provide a first empirical analysis of possessive constructions in Māori, the Indigenous language of Aotearoa New Zealand (ANZ). The A/O distinction in Māori presents a thorny grammatical issue for learners of the language but also for theoreticians trying to document its use. Contributing to these difficulties is the lack of large corpora of naturally occurring Māori, a problem which can be in part alleviated by harvesting Twitter posts. The article makes use of novel visualization techniques to flesh out patterns of categorical data (possessive O vs. possessive A), confirming that Twitter users tend to adhere to classical descriptions of possessive patterns found in grammars of the language. While the O form remains dominant, in keeping with grammar descriptions, users were shown to stray away from prescribed usage more often in cases where the A-possessive marker would have otherwise been expected, raising the possibility that a change may be underway (towards a single-form possessive).

A second grammatical analysis is given by Ruytenbeek, who also analyses Twitter posts, in a bid to make a connection between user stance and the choice of linguistic expressions encoding it. He argues that commuters posting evaluations of the French and Belgian national railway systems tend to frame their complaints using longer expressions (e.g., X is not positive-ADJ, *pas capable* 'not capable') rather than shorter ones (e.g., X is negative-ADJ, *incapable* 'incapable') in order to mitigate the face-threatening act of their complaint and to increase politeness, in the hope of minimizing offence.

O'Neill investigates reaction GIFS and images in a minoritized Community-of-Practice (COP) in Twitter. Complementing corpus linguistics methods (a body of GIFS and images collected from Twitter) with an online survey of users from the COP of interest, O'Neill reveals how non-traditional linguistic resources are enlisted by members of the community to create four distinct, albeit intersecting feminine-coded personae: the Sassy Queen, the Battle-Axe, the Flamboyant Queer and the Hun. The GIFS and images analysed are imbued with creativity and humour, recycling and remixing existing memes and cultural tropes. These elements both confirm users' affiliations within the respective COP, as well as further reinforcing constitutive links between linguistic resources and the COP.

Staying with the theme of humour, Schaefer investigates the use of memes and reels in German language youth radio posts on Instagram. In particular, the analysis focuses of the mixing of Anglicisms within German language posts by professional journalists managing online youth radio accounts. By teasing out notions of semantics and economy of expression, Schaefer investigates the lexical choices made by users on Instagram with a two-pronged approach: first, by collecting examples from Instagram to build a corpus, and second, by conducting semi-structured interviews with the users posting such content. The patterns uncovered show the strategic use of cross-linguistic resources (here, English words) to appeal and relate to a younger audience. In line with language contact research from other domains and languages, which argues for a social, symbolic role of loanwords (Zenner et al., 2019), this article draws on social media discourse to highlight the social role of Anglicisms in building the image and reputation of contemporary radio.

The final contribution is an analysis of the use of syntactically positive *anymore* in a large sample of English language tweets by Strelluf and Hills. The use of the adverb *anymore* is deemed ungrammatical in sentences that do not involve negative polarity (i.e., in those with positive polarity). For example, leaving out the negative particle *not* would render sentences like *He does not play for them anymore* ungrammatical in standard English. However, the occurrence of *anymore* in precisely these types of sentences, for instance *I refuse to lose in eight-ball anymore* and *I'm too stressed to care anymore*, persists in certain, albeit rare, and often regional contexts. This peripheral occurrence has made it difficult to acquire sufficient naturally occurring data to study this outside of experimental settings. Strelluf and Hills deploy sentiment analysis on a large body of Twitter posts containing syntactically positive *anymore* to show (among other things) the negative affect that accompanies this use. This is in line with findings uncovered in experiments on acceptability judgements from users. Their work expands on the general findings of such experiments by providing a more nuanced approach to negative affect characteristics of the construction and by showing regional differences in its interpretation. For example, syntactically positive *anymore* is associated with weaker language in the Eastern Midland region and with greater arousal approaching complaint-mode in the same region, but also in regions outside the Midland.

The collection of papers included here extends current knowledge of language on social media, and language more generally, often corroborating data collected in offline settings (interviews or experiments). Several themes emerge from this body of research. First, several articles illustrate the strategic and meticulous use of social media language. While it is certainly the case that not all SM language falls into this category, we see, for

example, that Māori language Tweeters use prescribed grammatical rules to form possessive constructions and that French-speaking commuters choose their adjective phrases in such a way as to increase politeness and minimize the offence caused by their negative evaluations.

Second, as also observed elsewhere (e.g., [Newton et al., 2022](#); [Shifman, 2014](#)), humour plays a major role on social media. In a world obsessed with attention and engagement, the ability to lure readers and increase visibility is paramount. Humour constitutes a great tool for achieving this. In keeping with the first theme, the use of humour is also strategic and reflects the linguistic repertoire of the audience which it is designed to address.

Third, this collection showcases the importance of tapping into sufficiently large datasets to be able to probe more peripheral corners of language (such as, A-possessive uses in Māori or syntactically positive *anymore* constructions in English) to increase our understanding of rare, but nevertheless revealing, language phenomena.

Despite the variety of languages analysed in the articles collated in this Special Issue, the range of SM platforms constitutes a limitation: examples are limited to Twitter and Instagram. It would be useful to probe data from other platforms, for example TikTok and YouTube, given their popularity. Another limitation of the work included here relates to the areas of linguistics that are represented. Although it is inspiring to see grammar, pragmatics, and variationist research, future work awaits morphological and phonological analyses of such data, especially given the multimodal nature of SM platforms.

Funding: This research received no external funding.

Acknowledgments: I thank the authors who chose to contribute to this Special Issue and to be part of the growing communities of scholars analysing social media language. Additionally, I thank the *Languages* editorial team for their help and support with the organisation of the articles included here.

Conflicts of Interest: The author declares no conflicts of interest.

Note

¹ This is not to imply that everyone has equal access to resources and social media content (see [Bender et al., 2021](#), among others).

References

- Androutopoulos, J., & Tereick, J. (2015). YouTube: Language and discourse practices in participatory culture. In A. Georgakopoulou, & T. Spilioti (Eds.), *The Routledge handbook of language and digital communication* (pp. 354–370). Routledge.
- Baron, N. S. (2008). *Always on: Language in an online and mobile world*. Oxford University Press.
- Bell, A. (2001). Back in style: Reworking audience design. In P. Eckert, & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 139–169). Cambridge University Press.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March 3–10). *On the dangers of stochastic parrots: Can language models be too big?* 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623), Virtual.
- Calude, A. S. (2023). *Linguistics of social media: An introduction*. Routledge.
- Crystal, D. (2001). *Language and the internet*. Cambridge University Press.
- Deumert, A. (2015). Linguistics and social media. In K. Allan (Ed.), *The Routledge handbook of linguistics* (pp. 561–573). Routledge.
- Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York University Press.
- Lee, C., & Barton, D. (2013). *Language online: Investigating digital texts and practices*. Routledge.
- Meltwater. (n.d.). *Global social media statistics*. Available online: <https://datareportal.com/social-media-users> (accessed on 11 May 2025).
- Miller, J. E., & Weinert, R. (1998). *Spontaneous spoken language: Syntax and discourse*. Oxford University Press.
- Newton, G., Zappavigna, M., Drysdale, K., & Newman, C. E. (2022). More than humor: Memes as bonding icons for belonging in donor-conceived people. *Social Media+Society*, 8(1), 1–13. [[CrossRef](#)]
- Page, R. (2012). The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse & Communication*, 6(2), 181–201.
- Seargeant, P., & Tagg, C. (2014). Introduction: The language of social media. In P. Seargeant, & C. Tagg (Eds.), *The language of social media: Identity and community on the internet* (pp. 1–20). Palgrave Macmillan.
- Shifman, L. (2014). *Memes in digital culture*. MIT Press.

- Statista. (2025). *Number of internet and social media users worldwide as of February 2025*. Available online: <https://www.statista.com/statistics/617136/digital-population-worldwide/> (accessed on 11 May 2025).
- Zappavigna, M. (2015). Searchable talk: The linguistic functions of hashtags. *Social Semiotics*, 25(3), 274–291. [CrossRef]
- Zenner, E., Rosseel, L., & Calude, A. S. (2019). The social meaning potential of loanwords: Empirical explorations of lexical borrowing as expression of (social) identity. *Ampersand*, 6, 100055. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.