(2025). *Discourse, Context and Media, 66*(August), 100931, special issue on 'Methodological synergies in the study of digital discourse,' edited by J. Androutsopoulos. https://doi.org/10.1016/j.dcm.2025.100931

Methodological Synergies in the Study of Digital Discourse: A Critical Reflection

Susan C. Herring
Indiana University, Bloomington

I. Introduction

This commentary offers a critical reflection on the articles featured in this article collection of *Discourse, Context, & Media* on 'Methodological Synergies in the Study of Digital Discourse.' Taken together, the articles examine the intersection of small-scale, contextualized, manual, sociolinguistic analysis and discourse analysis (henceforth: sociodiscourse analysis) with large-scale, automated, corpus and computational methodologies.

My comments are organized into two main parts, starting with a consideration of what constitutes methodological synergy in the domain of the article collection, broadly construed. I address this by providing a brief historical overview of synergistic approaches involving sociodiscourse analysis and computational methods and situating the present collection within this trajectory, highlighting what is familiar and what is novel about the articles in the collection. In the second half, I introduce a conceptual model grounded in Computer-Mediated Discourse Analysis (CMDA) (Herring, 2004) as an example of how synergy can be incorporated within a methodological paradigm focused on digital discourse. I propose that this model can serve as a guide for evaluating how successfully methodological synergy is achieved in studies that combine sociodiscourse and computational approaches, such as the articles in this collection. Specifically, I discuss two key criteria suggested by the model:

1. *To what extent* does the study integrate computational methods with sociodiscourse methods?
2. How *substantively* does each methodological component contribute to the analysis?

In addressing these questions, I advance recommendations for computational sociodiscourse analytic studies that aim to create meaningful methodological synergy. The commentary concludes by looking ahead to a future where generative AI and other advanced computational techniques are increasingly accessible, and by highlighting the opportunities and challenges they present for methodological innovation in sociodiscourse research.

II. Historical Context: The Emergence of Synergistic Methodologies

The concept of methodological synergy—particularly the integration of computational and qualitative approaches—has a well-established lineage in linguistics. Early instances can be traced to the rise of corpus linguistics, a subfield that emerged alongside the construction of large-scale computerized corpora, such as the Brown corpus of written American English[1] and

[1] For the Brown corpus, see Francis and Kučera (1964).

its British English counterpart, the London-Oslo-Bergen corpus,[2] in the 1960s and 1970s. The following decades saw the publication of foundational studies that applied corpus methods to sociolinguistic questions. Notable among these are D. Sankoff's introduction of the VARBRUL computer program for analyzing sociolinguistic variation in spoken corpora (Sankoff, 1975; Sankoff & Labov, 1979), Biber's (1986) multidimensional factor analysis of spoken and written registers, and Mair and Hundt's corpus-based investigations of real-time language change (e.g., Hundt & Mair, 1999; Mair, 1997, 2009). These studies were innovative for their time precisely because they employed computational tools to address questions traditionally situated within sociolinguistics.

The explosion of internet-based communication in the mid-1990s further catalyzed the development of corpus-based sociodiscourse analysis. Textual computer-mediated communication (CMC) was soon recognized as a rich source of data due to its accessibility, abundance, and pre-transcribed nature (Herring, 1996). There followed a wave of studies examining digital discourse across diverse sociotechnical platforms, languages, and contexts (e.g., Androutsopoulos, 2003, 2006; Georgakopoulou, 1997; Herring, 1994, 1999, 2004; Lee, 2007; Paolillo, 2001; Rintel, 1997; Siebenhaar, 2006). While methodologically innovative in many respects, these studies did not necessarily exemplify methodological synergy. Methodological synergy, in my view, arises not merely from the use of online data, but from the integration of analytical tools and frameworks drawn from distinct disciplinary paradigms.

Around the same time, researchers began digitizing existing corpora and compiling new ones based on various CMC modalities. In Europe, significant efforts have been directed toward constructing large-scale, reusable CMC corpora (e.g., Beißwenger & Storrer, 2008; Dürscheid & Stark, 2011; Poudat & Landragin, 2017). Sociodiscourse analysts have developed project-specific CMC corpora and applied corpus linguistic methods to analyze them from sociodiscourse perspectives (e.g., Tagg, 2009; Yates, 1996). Relatedly, computational linguists have applied computational methods to address variationist sociolinguistic questions using existing social media data. This computational sociolinguistics approach (Nguyen et al., 2016) recalls the early sociolinguistic corpus studies mentioned above, but with application to CMC corpora. These latter studies exemplify methodological synergy not simply because they involve digital data, but because they integrate computational techniques within sociolinguistic and discourse-analytic frameworks.

III. Evaluating the Contribution of 'Digital Sociodiscourse Analysis'

In his editorial introduction, Jannis Androutsopoulos employs the term digital discourse analysis to characterize the article collection as a whole. Because of the considerable overlap between sociolinguistics and discourse analysis in the articles in the collection, however, I will refer to the overall approach as digital sociodiscourse analysis. To assess the unique contribution of this approach, it is important to first clarify the meaning of the label, as it could be interpreted in different ways. That is, digital sociodiscourse analysis could denote either: (a)

[2] For the London-Oslo-Bergen corpus, see Johansson, Leech, and Goodluck (1978).

the use of digital-computational methods to conduct sociolinguistic or discourse analysis, or (b) the sociolinguistic or discourse analysis of digital data.

Collectively, the articles in the collection represent both (a) and (b). However, as noted earlier, (b) is not necessarily synergistic. Moreover, (a) is not new; it dates back several decades. What, then, is the distinctive contribution of the present collection?

In my reading, what sets this collection apart is its explicit engagement with methodological synergy as both a conceptual and an analytical focus. The former is evident in the deep methodological reflections offered by Georgakopoulou and Androutsopoulos, as well as in the collection's attention to diverse forms of synergy: between large-scale and small-scale data (Androutsopoulos; Yudytska), qualitative and quantitative methods (König), interactional and distributional analyses (Ilbury), and discourse-pragmatic approaches and automated abusive language detection (Long & Kübler). These contributions exemplify a dual focus, as described in the introduction to the collection: They present empirical findings while simultaneously reflecting on the methodological frameworks that underpin them.

The empirical analyses themselves stand on their own merits, shedding light on previously understudied sociodiscourse phenomena. These include the use of punctuation and typography as ideological double voicing in online political discourse (Androutsopoulos), textual representations of pitch contour in WhatsApp messages (Ilbury), the sequential organization of text and audio messages in WhatsApp conversations (König), and the influence of device type on the use of structural CMC features (Yudytska). Relatedly, several of the articles introduce useful new terminology—e.g., "transmodal interaction" (König) and "platformed discourses" (Georgakopoulou)—to describe digital sociodiscourse phenomena.

Of course, the most basic way in which the studies in this collection differ from early computational sociolinguistic research is in their focus on born-digital data, drawn from platforms such as Twitter, Reddit, Instagram, WhatsApp, Discord, news websites, and text message exchanges. Sampling and interpreting these data require taking into account—in addition to their sociocultural contexts—the platforms' *affordances*: what behaviors the technology facilitates or, conversely, makes less likely. Traditional writing and speaking also involve technologies of production and transmission—paper and pen or typewriter for writing; the human articulatory apparatus and the air through which sound waves travel for speech[3]—but these are so taken-for-granted that they are rarely considered as factors shaping sociodiscourse in non-CMC studies. In contrast, the relative novelty of CMC renders the medium of communication salient. Several of the articles in the present collection engage substantively with medium influences. For example, modality plays a central role in König's analysis of conversation involving text and audio messages on WhatsApp, and Yudytska's study, which compares language produced on mobile phones and laptop computers, directly explores technological effects.

Overall, the articles in the special collection highlight the potential and the growing importance of incorporating computational methods in digital sociodiscourse research. But how synergistic

[3] On the mediated nature of face-to-face communication, see Hollan and Stornetta (1992).

are the studies in actuality? To address this question, it is useful to examine one established paradigm—Computer-Mediated Discourse Analysis (CMDA)—that envisions methodological synergy between manual and computational approaches. In the following sections, I discuss how CMDA can serve as a conceptual anchor for evaluating methodological synergy.

IV. Synergy in Computer-Mediated Discourse Analysis (CMDA)

Although primarily a qualitative discourse analyst, I have advocated increasingly for the integration of large-scale computational approaches into digital discourse analysis in recent years. This includes promoting these methods as a natural extension of Computer-Mediated Discourse Analysis (CMDA), the methodological paradigm I developed (Herring, 2004), which was originally grounded in manual analysis of small datasets.

My support for methods that I do not personally employ stems from three main considerations. First, for any paradigm to remain relevant, it must evolve in response to changing technological and epistemological landscapes. Second, the sheer volume of online discourse data invites, if not necessitates, computational methods capable of efficiently identifying patterns across large datasets. Third, automated approaches offer the potential to scale up not only in terms of data volume but also in representativeness and generalizability. They can also capture infrequent-yet-theoretically-interesting phenomena that might not appear in smaller-scale studies.

At the same time, large-scale automated approaches often sacrifice contextual richness and interpretive depth. Certain discourse phenomena—for example, the layered, multimodal interactions on video-sharing platforms like TikTok (e.g., Herring & Dainas, 2025)—resist large-scale automation due to their reliance on nuanced contextual cues. Automated methods should therefore complement, rather than replace, the close, contextualized analysis that remains central to CMDA. More broadly, studies conducted on a human scale allow researchers to develop intimate familiarity with their data, enabling insights that are unlikely to emerge in large-scale analyses.

These tensions are represented schematically in Figure 1. Large-scale, automated, ‘top-down’ methods are positioned at the upper end, and small-scale, manual, ‘bottom-up’ methods are positioned at the lower end of the diagram, with arrows pointing towards the center, where aspects of the two methodologies come together. Text on the left side of the diagram indicates strengths (+) of each approach, while text on the right side indicates weaknesses (-) of each, which are to be avoided.

As a set of linguistic discourse analysis methods adapted for digital data, CMDA would originally have been located towards the bottom of the diagram in Figure 1. Over time, it has moved up toward the center, adding computer-assisted tools such as LIWC (a dictionary-based program), AntConc (a concordancer), and web-based sentiment analysis platforms. It also includes VisualDTA, a tool I co-developed with a former computer science student (Kurtz & Herring, 2006), which provides a semi-automated means of visualizing topic development in

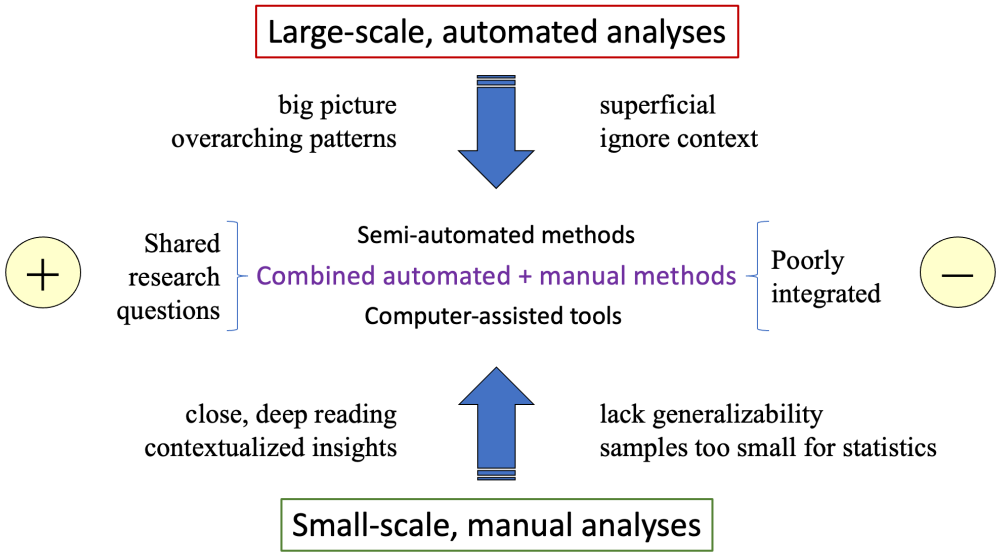

Figure 1. Potential methodological synergy in digital sociodiscourse analysis

computer-mediated conversations. As yet, the CMDA paradigm does not incorporate large-scale, automated methods, although the diagram suggests that it could. In fact, incorporating such methods could be essential for achieving true methodological synergy within CMDA.

V. Evaluating 'Methodological Synergy'

What does it mean for methodologies to be truly synergistic? The conceptual model in Figure 1 suggests two meta-methodological questions that could be asked to evaluate how successfully methodological synergy is achieved in CMDA and other sociodiscourse research.

Consider, first, the vertical axis. To what extent does the study integrate computational methods, be it in terms of research questions, theoretical assumptions, corpus construction, or analytical methods? A traditional sociodiscourse study that only applies an off-the-shelf computational tool such as a word frequency extractor, concordancer, or word cloud to count or visualize frequency distributions is not as synergistic as a study that uses a machine learning algorithm, for example, to classify sociodiscourse phenomena in a large corpus, together with manual interpretation of a sample of the results.

A second, related question is suggested by the horizontal axis in Figure 1. It is not enough to combine different methodologies in a single study; one should also ask: How substantively does each methodological component contribute to the analysis? One test is to ask whether a digital sociodiscourse study could stand on its own, with minimal modification, without its computational component. If so, the study is arguably not meaningfully synergistic. Ideally, the qualitative and computational methods should be well integrated and work in tandem to address a shared research question or questions. Furthermore, methodological choices should be driven by the research objectives, and not by the affordances of available tools or what methods the researcher is most familiar with.

Another way to think of these questions is that the first concerns *what* methods are included, while the second concerns *how well* they are integrated. The second question is about research quality, the overall methodological coherence of a study, while the first question is based on the assumption, in keeping with CMDA and the theme of the present collection, that integrating computational approaches in sociodiscourse analysis is an important direction to pursue in the current digital age. The guidelines suggested above are ideals, however. In practice, methodological synergy is a matter of degree, not of absolutes.

This brings us back to the question: How synergistic are the studies in this article collection on 'Methodological Synergies in the Study of Digital Discourse?' According to the vertical criterion, most of the studies are not optimally synergistic, in that they do not employ sophisticated computational techniques. This is understandable, since most sociodiscourse analysts (myself included) lack training in advanced computational methods. According to the horizontal (meaningful integration) criterion, some of the studies are well balanced between sociodiscourse and computational/corpus methods (Androutsopoulos, Yudytska), while others use computational tools only to extract data for analysis from larger corpora (Ilbury, König). Indeed, if I have one criticism of the collection, it is that the included articles are inconsistent in the extent to which they demonstrate methodological synergy between sociodiscourse and digital-computational approaches. In addition to the inconsistencies just noted, the article by Georgakopoulou differs from the others in that it is a methodological meditation on the author's research trajectory rather than a single synergistic study. Finally, Long and Kübler's contribution is farthest from the article collection's theme: Although they analyze disagreements in annotator interpretations of abusive language from sociodiscourse perspectives with the ultimate goal of improving automated detection, no automation is used in the study, nor does the word synergy appear in the article. To bring these diverse studies together under a single framework, the concept of methodological synergy must be interpreted broadly—as Androutsopoulos does in his introduction to the collection.

VI. Contemporary Developments and Future Directions

The boundaries between top-down computational approaches and bottom-up sociodiscourse analysis are increasingly porous. Research paradigms that once appeared as static binaries—manual vs. automated, contextualized vs. generalizable, interactional vs. distributional—are converging in mixed-methods studies such as those in the present collection that blur these distinctions.

Additionally, recent advances in machine learning and artificial intelligence have significantly expanded the methodological toolkit available to computational sociolinguists and discourse analysts. Large-scale studies such as Danescu-Niculescu-Mizil et al. (2013) on linguistic politeness in online forums, Abulaish et al. (2020) on figurative language detection in social media, and Jin (2022) on the identification of complaints and bragging speech acts in social media exemplify how computational techniques can illuminate enduring questions about social behavior through online discourse. These studies represent a form of methodological synergy that originates in the computational sciences and reaches toward the social sciences.

At the same time, large language models (LLMs) such as ChatGPT are making computational operations increasingly accessible to non-computer scientists. User-friendly prompt-based interfaces take requests in ordinary language, rather than requiring input via a programming language. LLMs can also identify sociodiscourse phenomena such as politeness or speech acts in data provided by the researcher, and they can be instructed to take context into account. While the quality of the interpretations produced by LLMs so far does not match the sophistication of interpretations produced by human sociodiscourse analysts, the growing capability and accessibility of LLMs holds significant promise for digital sociodiscourse research, with the potential to further dissolve disciplinary boundaries. Yet, despite this potential, few sociodiscourse studies have so far incorporated generative AI methods, nor do any of the articles in the present collection do so. Given the current enthusiasm surrounding generative AI, however, such research is likely on the horizon.

VII. Conclusion

This commentary has traced the evolving landscape of methodological synergy in digital discourse research, from early corpus-based sociolinguistic studies to the varied, mixed-methods approaches showcased in this article collection. I argued that the articles in the collection make a distinctive, valuable contribution to this landscape through their nuanced engagement with methodological synergy as a conceptual construct, and through their exploration of novel data types, underexamined communicative phenomena, and the complexities of digital communication.

I further critiqued the concept of methodological synergy by drawing on the CMDA paradigm to illustrate both the tensions and the possibilities it entails for computational sociodiscourse research. CMDA provides an aspirational blueprint for how traditional discourse frameworks can be adapted to meet the demands of large-scale digital data. It also proposes evaluative criteria for achieving optimal methodological synergy: the use of sophisticated computational methods, integrated in ways that are both substantive and analytically coherent. While neither the articles in this collection nor current CMDA research fully meet these ideal standards, the goal is increasingly attainable. This is thanks, especially, to advances in artificial intelligence, particularly the emergence of large language models, which democratize access to computational power.

It seems inevitable that at least some sociodiscourse researchers will begin experimenting with the integration of AI methods. It is even conceivable that a new paradigm of AI-augmented sociodiscourse analysis will emerge in the future. In either scenario, the central challenge will be to leverage the power of algorithm-driven techniques while maintaining the interpretive depth and insight that define high-quality sociodiscourse research. In this spirit, I encourage researchers to continue experimenting and critically reflecting on their methodological choices, so that synergy between qualitative and computational approaches becomes a meaningful engine of scholarly insight.

References

Abulaish, M., Kamal, A., & Zaki, M. J. (2020). A survey of figurative language and its computational detection in online social networks. *ACM Transactions on the Web (TWEB)*, *14*(1), 1-52.

Androutsopoulos, J. K. (2003). Online-Gemeinschaften und Sprachvariation. Soziolinguistische Perspektiven auf Sprache im Internet [Online communities and language variation. Sociolinguistic perspectives on language in the internet]. *Zeitschrift für germanistische Linguistik, 31*(2), 173-197.

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, *10*(4), 419-438.

Beißwenger, M., & Storrer, A. (2008). Corpora of computer-mediated communication. In A. Lûdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook.* Mouton de Gruyter.

Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language, 62*, 384-414.

Chanier, T., Poudat, C., Sagot, B., Antoniadis, G., Wigham, C. R., Hriba, L., ... & Seddah, D. (2014). The CoMeRe corpus for French: Structuring and annotating heterogeneous CMC genres. *Journal for Language Technology and Computational Linguistics*, *29*(2), 1-30.

Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., & Potts, C. (2013). A computational approach to politeness with application to social factors. *Arxiv Preprint Arxiv:1306.6078*.

Dürscheid, C., & Stark, E. (2011). SMS4science: An international corpus-based texting project and the specific challenges for multilingual Switzerland. In C. Thurlow & K. Mroczek (Eds.), *Digital Discourse: Language in the new media* (pp. 299-320). Oxford University Press.

Francis, W. N., & Kučera, H. (1964). *Manual of Information to Accompany a Standard Sample of Present-day Edited American English, for Use with Digital Computers.* Original ed. 1964, revised 1971, revised and augmented 1979. Department of Linguistics, Brown University.

Georgakopoulou, A. (1997). Self-presentation and interactional alliances in e-mail discourse: The style-and code-switches of Greek messages. *International Journal of Applied Linguistics*, *7*(2), 141-164.

Herring, S. C. (1994). Politeness in computer culture: Why women thank and men flame. In *Cultural Performances: Proceedings of the Third Berkeley Women and Language Conference* (p. 278-294). Berkeley Women and Language Group, University of California.

Herring, S. C. (1996). Introduction. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives* (pp. 1-10). John Benjamins.

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

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 learning* (pp. 338-376). Cambridge University Press.

Herring, S. C., & Dainas, A. R. (2025). Improbable conversations: Interactional dynamics in TikTok duets. *Discourse, Context, & Media, 63*, February 2025, 100821. Special issue, cc.

Hollan, J., & Stornetta, S. (1992, June). Beyond being there. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 119-125).

Hundt, M., & Mair, C. (1999). "Agile" and" uptight" genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics*, *4*(2), 221-242.

Jin, M. (2022). *A computational study of speech acts in social media*. Doctoral dissertation, University of Sheffield.

Johansson, S., Leech, G., & Goodluck, H. (1978). *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Department of English, University of Oslo.

Lee, C. K. (2007). Linguistic features of email and ICQ instant messaging in Hong Kong. In B. Danet & S. C. Herring (Eds.), *The multilingual Internet: Language, culture and communication online* (pp. 185-208). Oxford University Press.

Mair, C. (1997). Parallel corpora: A real-time approach to the study of language change in progress. In M. Ljung (Ed.), *Corpus-based studies in English* (pp. 195-209). Rodopi.

Mair, C. (2009). Corpus linguistics meets sociolinguistics: The role of corpus evidence in the study of sociolinguistic variation and change. *Language and Computers, 69*(1), 7-32.

Nguyen, D., Doğruöz, A. S., Rosé, C. P., deJong, F. (2016). Computational sociolinguistics: A survey. *Computational Linguistics, 42*(3), 537-593.

Paolillo, J. C. (2001). Language variation on Internet Relay Chat: A social network approach. *Journal of Sociolinguistics*, *5*(2), 180-213.

Poudat, C., & Landragin, F. (2017). *Explorer un corpus textuel: Méthodes-pratiques-outils*. De Boeck Superieur.

Rintel, E. S., & Pittam, J. (1997). Strangers in a strange land: Interaction management on Internet Relay Chat. *Human Communication Research*, *23*(4), 507-534.

Sankoff, D. (1975). VARBRUL version 2. Unpublished program and documentation.

Sankoff, D., & Labov, W. (1979). On the uses of variable rules. *Language in Society*, *8*(2-3), 189-222.

Siebenhaar, B. (2006). Code choice and code-switching in Swiss-German Internet Relay Chat rooms. *Journal of Sociolinguistics*, *10*(4), 481-506.

Tagg, C. (2009). *A corpus linguistics study of SMS text messaging*. Doctoral dissertation, University of Birmingham.

Yates, S. J. (1996). Oral and written linguistic aspects of computer conferencing. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives* (pp. 29-46). John Benjamins.