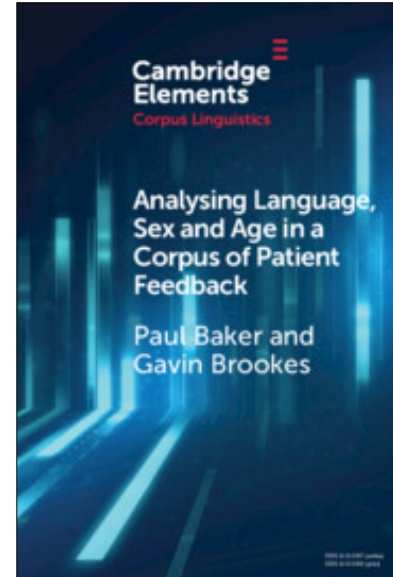


Paul Baker and Gavin Brookes, **Analysing Language, Sex and Age in a Corpus of Patient Feedback: A Comparison of Approaches**, Cambridge, UK: Cambridge University Press, 2022, 88 pp., \$20.00 (paperback).

Reviewed by
Qiuying Zhao¹
Xi'an International Studies University, China

After decades of development, language research based on electronic corpus has gradually become mainstream and led to the extensive integration of corpus research and adjacent disciplines. For instance, teachers and researchers who work in linguistic fields like applied linguistics, sociolinguistics, and discourse studies have increasingly used methods from corpus linguistics to unveil the relationship between language use and social identity. Corpus linguistics is a collection of methods and a research field concerned with analyzing language use based on large collections of naturally occurring language data (McEnery & Wilson, 2001). It continues to become increasingly complex, both in its methods and theoretical concepts. Authors Paul Baker and Gavin Brookes's ***Analysing Language, Sex and Age in a Corpus of Patient Feedback: A Comparison of Approaches*** is designed to meet the needs of researchers, teachers, and students who want to keep up with the changing field.



The book attempts to describe social identity in language use with a specialized corpus of patients' written feedback on cancer care services in England by comparing two broad approaches (using sociodemographic metadata and adopting a more data-driven approach). It explores the relationship between discourse and social identity in depth, which makes great contributions to the "social turn" of discourse analysis and provides important implications for the vigorous development of corpus linguistics.

The book is divided into four chapters and describes two case studies that each compare two approaches to social identity variables (sex and age) in a large corpus of 14,403,649 patients who provided feedback on the National Health Service (NHS) cancer care services in the United Kingdom. Additionally, two methods are presented for each case study, the first utilizing sociodemographic metadata tags to create two subcorpora according to patients' sex or age, and the second categorizing cases in which patients verbally reference their sex or age in their comments. Finally, the authors critically discuss the contrasting findings from both approaches, which brings new insight into key corpus-aided studies of language and society.

Chapter 1 consists of five subsections. Following a brief introduction of the corpus methods and the book's objective, the authors define sex, gender, and age and make a clear distinction between sex and gender. Thereafter, the authors address the conception of identity, how it relates to language use, and a mainly

¹This research was supported by National Social Science Fund of China (No. 18AYY006).

theoretical perspective on their relationship. Additionally, the authors examine existing approaches to language and identity in corpus studies, emphasizing that sociodemographic metadata annotations are typically used to group language users. In the final subsection, the authors describe the data and methods, as well as the techniques, of corpus linguistics, such as keywords, collocation, and concordancing, which are used in the analysis of the following two sections. For example, the top five keywords for male patients' comments are "class," "bladder," "treatment," "good," and "hospital" (p. 18), while the top five keywords for female patients' comments are "I," "kind," "felt," "n't," and "amazing" (p. 27).

Chapter 2 describes the first case study, which investigates the possible impact of patient sex and discourse surrounding gender on feedback from cancer care services. To pursue this aim, the authors apply two approaches, both based on the keyword technique, with one treated through the keywords appearing in the two examples, and the other considered through the keywords specific to either sample. In the first approach, the authors use the sociodemographic metadata tags with the patients' feedback to compare comments from males and females. Some keywords generated from this comparison indicate differences in the thematic content of the patients' comments and what is centered on or highlighted in them. For instance, male patients focused on the transactional aspects of care, such as operations, while female patients often concentrated on the individuals performing those operations and on the relationships they built with the staff. In the second approach, the authors demonstrate corpus-internal indications of patient identity, in which they base their analysis on the comments where patients verbally reference their sex identity. The authors also elaborate on the advantages and disadvantages of both approaches and note that the two approaches offer complementary insights into the influence of patient sex identity and gendered discourse. Furthermore, they determine that the first approach is much more appropriate to use, especially when the researchers could build a fairly large corpus.

In chapter 3, Baker and Brookes consider the second case study on how patient age relates to positive and negative feedback. This chapter takes a different approach with the sample and corpus analyses rather than replicating the method from chapter 2. First, the authors conduct a qualitative analysis grounded on closely reading small samples of patients' feedback and categorizing positive and negative ones, such as "Oncology staff work well together" (p. 48) and "Got up and left to sit down when I fainted alone on the floor" (p. 49). Thereafter, they conduct corpus analysis of all the patients' feedback from the two age groups and examine collocates of the most frequent evaluative words, which also represent positive and negative feedback. Next, they conclude their findings: Both age groups are likely to give positive evaluations related to staff, care, and treatment; few differences between age groups are found, with staff making up about 40% of negative evaluations, followed by information, care, and waiting times. Finally, the authors compare the qualitative and corpus approaches to the age groups, which reveals more similarities than differences, consider their respective pros and cons, and raise several questions for further study, such as the question about what it really reveals regarding different complaints about staff and communication.

In chapter 4, the final chapter, the authors further compare findings from both approaches, noting their advantages and limitations respectively, and attempt to outline how the analysis could have been optimized. The improvement of research methods will inform future development trends of discursive identity constructions. At the same time, the authors state that in a large corpus of texts, the appropriate approach for researchers engaged in studying the relationship between language and identity depends on the level of granularity or accuracy required and the data. They recommend combining these two approaches when it is

necessary and look forward to future development by focusing on the relationship between identity and linguistic evaluation, which is also an understudied topic.

Baker and Brookes' groundbreaking book on language and identity is a valuable addition to the existing literature. It highlights language to use with advances in sociodemographic information and corpus-aided studies. The book is highly systematic, not only introducing major concepts but also contributing profound commentary and in-depth explanations regarding two different approaches and research methods of great importance. For instance, combining the annotation-based approach with searches for identity citations and complementing a whole-corpus approach with that of working closely with samples will likely fill research gaps in this field and provide a richer description of the data (p. 75). It is highly recommended for its academic, cutting-edge, and pioneering role.

Given the focus, this book offers strong guidance for researchers and postgraduates in foreign language applied linguistics and discourse studies; it also serves as a theoretical reference and is practical for corpus linguistics teaching. It can be used as a textbook or reference for graduate students majoring in linguistics and applied linguistics. Simultaneously, the analysis of patient feedback is helpful and meaningful for those interpreting such feedback, providing appropriate recommendations for improvements. The guidelines are especially valuable for scholars engaged in sociolinguistics, corpus linguistics, and discourse studies.

Regarding the methodology, this book concentrates on corpus-aided quantitative and qualitative methods. Such an analysis method has obvious advantages at the level of language ontology, such as vocabulary, phrases, structure, and grammar, but it may have limitations at the macro-level context and sociocultural levels. To make up for the insufficiency of corpus-based discourse analysis and to expand the field of discourse analysis, some scholars (Hamdi, 2022; Lee, 2019) have attempted to apply text mining techniques to discourse analysis. Text mining is a comprehensive technique involving text classification, text clustering, topic modeling, sentiment analysis, and statistical methods for mining knowledge from unstructured text. Therefore, if the authors were to add another chapter on patients' feedback using text mining, that is, to incorporate the theories of discourse analysis with quantitative text mining techniques to effectively triangulate their findings, it would further enrich the case studies of discourse analysis.

References

- Hamdi, S. (2022). Mining ideological discourse on Twitter: The case of extremism in Arabic. *Discourse & Communication*, 16(1), 76–92. doi:10.1177/17504813211043706
- Lee, C. (2019). How are "immigrant workers" represented in Korean news reporting?—A text mining approach to critical discourse analysis. *Digital Scholarship in the Humanities*, 34(1), 82–99. doi:10.1093/llc/fqy017
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction* (2nd ed.). Edinburgh, UK: Edinburgh University Press.