

doi:[10.1017/S136067432100023X](https://doi.org/10.1017/S136067432100023X)

Paula Rautionaho, Arja Nurmi and Juhani Klemola (eds.), *Corpora and the changing society: Studies in the evolution of English*. Amsterdam and Philadelphia: John Benjamins, 2020. Pp. xii+305. ISBN 9789027205438.

Reviewed by Naomi Nagy, University of Toronto

This collection of papers presented at ICAME 39 (Tampere, Finland, 30 May – 3 June 2018) targets an innovative use of corpora: the exploration of how society has changed over the time period covered by the corpus, querying whether (and how) we can distinguish linguistic change from societal change. I must begin with the disclaimer that I am a sociolinguist and not a corpus linguist. (The contrast between variationist sociolinguistics and corpus linguists in approaches to historical linguistics is described in Szmrecsanyi 2016.) This, perhaps, allows me to particularly appreciate that this volume brings new material to the table for consideration by sociolinguists. On a slightly less positive note, for a variationist sociolinguist accustomed to considering statistically supported comparisons based on predictive multivariate models, rather than comparing raw rates (that may be highly influenced by distributional differences between samples being compared), these papers presented some unsatisfying reports of results which compare raw numbers of collocations, even though the sample size (by time period and/or genre) differs, sometimes by a large degree. While it is not easy to determine an appropriate ‘envelope of variation’ in some cases, this distinction between corpus linguistics and (most) variationist sociolinguistic studies could benefit from further cross-pollination. This volume can serve as an important step toward such work.

These reservations notwithstanding, the topic intrigued me to the extent of incorporating it into a graduate seminar in sociolinguistics in 2021. Some of what I write here was influenced by productive discussions with students in that course: Anissa Baird, Nicki Butler, Vidhya Elango, Christopher Legerme and Justin Leung, to whom I extend thanks.

As ICAME’s focus is English machine-readable texts, only the English-speaking world is examined. But the geographic focus is, in fact, much narrower: only Britain, primarily England, and, in four chapters, the United States, contribute data to this book. A wide range of timeframes, however, is included, spanning from Old English to Present-day English, and the corpus types include both specialized and general collections and spoken and written texts. Conveniently, the software and corpora used in each chapter are specified in a dedicated section of the chapter.

The two parts of the book (part I ‘Changing society’, part II ‘Changing language’) do not engage with each other, so I will discuss them separately. While I was disappointed at the lack of interchange between these themes, each half of the book, on its own, contains well-written chapters. These present new findings and clear descriptions of the methods applied (if not always clear justification of the methods). Due to space considerations, I am not able to comment on every contribution.

Part I

The highlight of the book is the first chapter, by Martin Hilpert. He elegantly teases apart linguistic and social change by considering whether the ‘diminishing social value of interpersonal authority is reflected in changing patterns of language use’ (p. 3), specifically changes in animacy and verb selection in *make*-causatives (‘A makes B do something’). He illustrates first how some previous attempts to illustrate this do not use reliable methods. For example, Greenfield (2013) looked at lexical frequencies of words that we might expect to decrease in frequency when interpersonal authority decreases, e.g. *authority*, *obedience*, vs some that might increase, e.g. *individual*, *choose*. Counts from Google Books corpora illustrate these frequency changes, but Hilpert notes (as did Liberman 2013) that the evidence is not conclusive for several reasons. These relate primarily to disconnects between frequencies of words and of their denotations, polysemy, spurious correlations in large datasets and appropriate statistical testing, particularly considering diachronic alignment (p. 6).

That individual lexical items (e.g. *authority*, *individual*) behave less systematically than grammatical constructions (e.g. *make*-causatives) will be unsurprising to variationist sociolinguistic readers. It is frequently noted that lexical items contrast with grammatical patterns, in terms of systematic patterning. This has been attributed to their higher level of accessibility (Sankoff 2002; Walker 2010: 5; Ravindranath Abtahian & Kasstan 2020) and their (relatively) fixed association to (truth-conditional, qualia-based, as opposed to social) meaning (Labov 2010).

Hilpert outlines five ‘pitfalls’ to avoid in seeking evidence for social change in diachronic linguistic variation (section 2) and illustrates how to avoid them via an illustrative analysis of *make*-causatives (section 3). The argumentation is clear and well supported by examples and graphs, making it easy to accept his conclusion that there is not strong evidence to support ‘the hypothesis that social change, in the form of waning interpersonal authority, [is] visible in the changing use of a grammaticalized construction that serves the purpose of expressing ... that kind of authority’ (p. 25). More broadly, he cautions that we should be sure to have non-linguistic, in addition to linguistic, evidence, to support claims of societal change, and to incorporate inferential statistical tests in our analyses. Hilpert’s cautions are, unfortunately, only partially evident in other chapters.

Gerold Schneider addresses ‘how societal and linguistic changes can be detected’ (p. 29), with a case study examining the vocabulary of poverty and the industrial revolution. He begins by establishing language-external data about (male) English soldiers’ height as a ‘gold-standard’ measure of poverty. (This is a measure whose validity, without checking the cited sources, I question. Average height could change if recruiting strategies change, if the ethnic make-up of the military changes, if tall people emigrate and don’t become soldiers ...) The key is that there is an external measure, to which the author seeks correlation with linguistic variation to validate its indexing of poverty in a corpus of texts.

Illustrating a well-justified corpus-driven approach (in which the hypotheses arise from the data, rather than being motivated by external theoretical assumptions, p. 35), the chapter evaluates several methods of seeking connections between the societal change and the linguistic: *document classification*, *topic modelling* and *conceptual maps*. Document classification fails because it is too coarse-grained: it seeks to group texts by (unstructured) lists of lexical items they have in common, without attention to the sense of the word and suffering from fluctuations caused by sparse data in certain timespans. The refinement of topic modelling, which looks at co-occurrence patterns of words within sections of a document, fairs slightly better, as co-occurrence patterns reveal semantic nuances. Finally, conceptual maps, or kernel density estimation, also explore the distributional semantics, this time via patterns of co-occurrence across texts. Issues including unbalanced corpora, spelling variants (especially those that drift diachronically), OCR issues arising from text-digitization, and tokenization errors are noted which, in tandem with large corpus sizes, necessitate tools for mediation. Useful tools are presented though not explained at a level to be useful to those not already familiar with them.

The link between societal change and linguistic change is method-external, in that reports of military records were used to establish the rises and falls of poverty, and then corpus-analysis tools were tested to see whether they could provide fluctuations that matched patterns in that measure. No statistical methods of determining how good the match is are noted, but the chapter concludes with interesting comments about the types of sentiments about poverty that were uncovered.

Like Schneider's chapter, Maura Ratia's main aim is to show corpus evidence to support a non-linguistically documented societal change. The chapter traces differences in how patients were viewed in medical texts over three centuries, concluding, via collocation analysis, that there was a shift from the patient as an 'object' of treatments to the patient as an experiencer. That is, the patient gained some humanity, in the view of medical experts, by 1800, or, in other words, the texts shifted from describing treatments to describing patients. The chapter succinctly provides details on methods, but suffers from some lack of diachronically continuous data samples. It is not convincing that the conclusion of change is justified, based on the lack of inferential statistical evidence of diachronic differences. That is, the 'top ten' collocates differ between the two time periods compared but no information about period-internal comparison is provided (as suggested by Hilpert), nor are statistical tests used to see whether frequencies (or Mutual Information scores) are robustly different between periods.

It's not clear how this finding could be useful from a linguistic perspective, in spite of its historical import. This is, perhaps, due to one aspect that could be better presented: the examples. They are plentiful, and they illustrate a range of meanings, but they are, seemingly, not organized into clear hierarchies or patterns; they do not seem to support generalization. Information about the methods of analysis is sparse. It is noted that selecting texts from different genres, and from different authors, muddies the water but is, of course, unavoidable in exploring a 300-year time span.

Gavin Brookes and David Wright offer a strong contribution in terms of methodology, illustrating a principled ‘corpus-assisted approach to Critical Discourse Analysis’ (p. 113). The typical traits of quantitative analysis (lots of data, little in-depth analysis of individual tokens), on the one hand, and qualitative (necessarily smaller data sets, in-depth if not as systematically organized analysis of individual tokens) on the other, are balanced and held in tension by this combined approach. The authors first determined the most frequent collocates of ‘speak English’ (in titles and lead paragraphs of newspaper articles) and then examined attitudes represented by a subset of the relevant sentences. These were grouped into categories relating to *learning English* (where multilinguals are criticized both for doing too well in school and for not doing well enough), *proficiency*, *multilingualism* and (the cost of) *services*. A clear view emerges of the bias in the category of media selected for examination, well captured in the chapter title, ‘From burden to threat’. This chapter provides valuable insight for those interested in combatting such bias.

The chapter presents an appealing query (how have representations of non-native English speakers changed in the popular press?) and replicable methods of exploring changing collocation patterns. Methods are clear and invite replication, perhaps because the authors were, indeed, replicating and expanding a previous study. No attention is given, however, to issues of linguistic change, only societal. Given that focus, Elango (2021: 5) notes that they could have better connected their findings to ‘recent literature on neoliberalism, language, and migration (cf. Allan & McElhinny, 2017, pp. 85–9) [reporting on immigrants being expected to manage their own linguistic integration rather than receiving government support], and literature on neoliberalism and language more generally (Heller & Duchêne, 2016; Heller & McElhinny, 2017)’. Otherwise, the specifics of the language ideologies to which the authors make reference are unclear.

Part II

This second half of the book, with a focus on linguistic, rather than societal change, includes three chapters on the grammaticalization of intensifiers (Aijmer, Blanco-Suárez, Schweinberger). My review focuses on this cohesive group. In this part of the book, there is repeated reference to ‘grammaticalization’ across the chapters, but there is no clear definition or apparent consistency in the meaning of the term. Brinton’s contribution provides a ‘big picture’ of the criteria that constitute the grammaticalization process, while other chapters have more focused and, in some cases, apparently contradictory criteria. To illustrate, Karin Aijmer, Zeltia Blanco-Suárez and Martin Schweinberger provide extensive and partly overlapping introductions to the concept of grammaticalization. The overlaps could have been mitigated through editorial oversight, while the coverage provided still left my students and me a bit unclear on what the patterns analyzed as grammaticalization share. We were left wondering what, exactly, is intended by grammaticalization in corpus linguistics, particularly given that, in at least one contribution, the term is used to

describe a process of change that did not stem from a lexical item, according to available evidence. Specifically, Laurel Brinton notes that her exploration of *that is not to say* is ‘not an entirely prototypical case of grammaticalization as certain parameters ... are inconclusive and there are no “lexical” uses of the form’ (p. 251). We also wonder: is there a specific path that can serve as a diagnostic of this type of change? What criteria are used to establish a structure at the starting point of the grammaticalization process? Can one ever establish an endpoint of the process empirically?

Aijmer investigates *absolutely*, illustrating that its increase in frequency of use is related to its shift from a degree modifier to an emphaser and then to a discourse marker. This illustrates the grammaticalization path (in a two-decade timespan, if I understand the corpus design correctly) involving increasing subjectivity. Reading from a variationist perspective, the absence of an ‘envelope of variation’ or ‘denominator’ (e.g. all intensifiers) is troubling. One might see the data presented as evidence of a rise in the use of a particular intensifier, as the author does, but it might equally reflect a rise in the number of intensification-worthy items people (in the relevant groups) see fit to discuss. That is, there is no consideration given to whether this intensifier is replacing another (or others).

One interesting feature of the comparison between the two corpora representing different time periods is that the frequency of *absolutely* drops (from BNC1994 to BNC2014) for the 15- to 24-year-olds but rises for all other groups. As that group includes the adolescent period where we anticipate innovation, it seems likely that this pattern indicates that these speakers are already off and running with a newer intensifier. Interestingly, this change is carried mostly by the male speakers – the rate for females’ use of *absolutely* is almost the same in the two time periods. However, with no defined envelope of variation for this exploration, nor query of other intensifiers, we cannot confirm these findings. Detailed comparisons with other studies of this intensifier suffer the same shortcoming. This could well be the source of some of the variation in accounts of how intensifiers change (e.g. the comparison of the roles of males and females, pp. 148–9). In fact, the end of that section, which looks at the different roles of intensifiers, and who favours which, really motivates the need for a more constrained approach. Particularly the comparison of uses of the intensifier with different heads, between BNC1994 and BNC2004, invites such an approach. As I look at it, I wonder how the differences in use of *absolutely* with adjectives vs response markers, for example, line up with differences in use between age groups, or between males and females. Comparisons are made, but without attention to statistical significance (pp. 150–6). The most intriguing finding is apparent-time evidence (comparison of age groups sampled in one time period; cf. Bailey *et al.* 1991), suggesting change in the opposite direction from the real-time comparison (figures 2 and 3).

Similarly, Blanco-Suárez presents a study of the grammaticalization of two intensifiers, *deathly* and *mortal*, without consideration of their role in a larger intensifier system. She explores a tantalizing hypothesis: that more positive and neutral collocations over time could serve as evidence of the grammaticalization of these ‘inherently negative’ intensifiers (p. 172). Given the methodology, we have to weigh that against another

interpretation, one that might reflect societal change rather than linguistic: the whole discourse (the whole society?) is becoming more positive (or neutral) over time. We also wonder if genre was controlled as it is noted that additional data from a collection of novels were used for the LModE period but not the EModE (p. 173).

The discussion in section 4 could be better organized, perhaps by theme or variable, rather than presented in chronological chunks, as that information is reprised in the figures in section 5. The figures, incidentally, would be more comparable if the y-axis scale were held constant, if the figures meant to be compared were presented adjacent to each other, and if the legend were correct in figure 3.

After these two studies that examine isolated elements of the intensifier category (*mortal*, *deathly* and *absolutely*), it is helpful to see the chapter by Schweinberger, which examines a much larger set of intensifiers. As his data come from American corpora, this chapter may not be directly comparable to the previous chapters, which focused on British English. Methodologically, however, it forms an important complement. This study is restricted to the subset of intensifiers called ‘amplifiers’, which should include words like *mortal*, *deathly* and *absolutely* (though only the latter appears in the results in figure 1 (p. 231)). In contrast to the two studies just discussed, Schweinberger considers ‘all variants that occur in the variable context’ (p. 224), making it possible to measure the rise of one lexical item against its competitors, rather than in the absolute, as well as considering the conditioning or interdependency of intensifiers on particular adjectives. This avoids the pitfalls noted for the two other contributions on intensifiers.

This chapter also contains a thorough review of studies of amplifiers. Detailed statistical analysis, of several types, is explored, well-explained, and accompanied by tight argumentation, allowing this chapter to serve as a valuable model for other explorations. The author notes differences in the factual outcomes compared to other studies, connecting them to differences in methods (using fiction texts rather than conversational speech). While he notes that ‘fiction data is more spoken- or speech-like compared with more formal written text types, it remains a written rather than a spoken genre’ (p. 242). Perhaps more important is that fact that only some (unspecified) portion of a work of fiction is normally ‘speech-like’, that is, quoted dialogue.

Overall, this collection of eleven chapters is well written and well edited, though it is best considered as two halves that do not form a cohesive whole. In this way, they provide intriguing motivation to seek further ways to disentangle the changes that are ongoing in society and language, as well as clear methodological models.

Reviewer's address:

Department of Linguistics

University of Toronto

Sidney Smith Hall, 4th floor

100 St George St

Toronto, ON M5S 3G3

Canada

naomi.nagy@utoronto.ca

References

- Allan, Kory & Bonnie McElhinny. 2017. Neoliberalism, language, and migration. In Suresh Canagarajah (ed.), *The Routledge handbook of migration and language*, 79–101. Abingdon: Routledge.
- Bailey, Guy, Tom Wikle, Jan Tillery & Lori Sand. 1991. The apparent time construct. *Language Variation and Change* 3(3), 241–65.
- Elango, Vidhya. 2021. Critical review of Brookes & Wright. University of Toronto course paper.
- Greenfield, Patricia M. 2013. The changing psychology of culture from 1800 through 2000. *Psychological Science* 24, 1722–31.
- Heller, Monica & Alexandre Duchêne. 2016. Treating language as an economic resource: Discourse, data and debate. In Nikolas Coupland (ed.), *Sociolinguistics: Theoretical debates*, 139–56. Cambridge: Cambridge University Press.
- Heller, Monica & Bonnie McElhinny. 2017. *Language, capitalism, colonialism: Toward a critical history*. Toronto: University of Toronto Press.
- Labov, William. 2010. *Principles of linguistic change: Cognitive and cultural factors*, vol. 3. Oxford: Wiley-Blackwell.
- Liberman, Mark. 2013. The culturomic psychology of urbanization. *Language Log*, 18 August 2013. <http://languagelog ldc.upenn.edu/nll/?p=5985> (accessed 31 December 2018).
- Ravindranath Abtahian, Maya & Jonathan R. Kasstan. 2020. Language contact and sociolinguistic variation. In Raymond Hickey (ed.), *The handbook of language contact*, 2nd edn, 221–39. Oxford: Wiley-Blackwell.
- Sankoff, Gillian. 2002. Linguistic outcomes of language contact. In J. K. Chambers, Peter Trudgill, & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*, 638–68. Oxford: Blackwell.
- Szmrecsanyi, Benedikt. 2016. About text frequencies in historical linguistics: Disentangling environmental and grammatical change. *Corpus Linguistics and Linguistic Theory* 12(1), 153–71.
- Walker, James A. 2010. *Variation in linguistic systems*. Abingdon: Routledge.

(Received 9 June 2021)