

Paula Rautionaho, Arja Nurmi and Juhani Klemola (eds.). *Corpora and the changing society. Studies in the evolution of English* (Studies in Corpus Linguistics 96). Amsterdam and Philadelphia: John Benjamins Publishing Company, 2020. 305 pp. ISBN 9789027205438 (HB).

Review

Daniela Pettersson-Traba\*

Complutense University of Madrid

The volume under review, edited by **Paula Rautionaho**, **Arja Nurmi**, and **Juhani Klemola**, presents selected papers from the *39th ICAME conference*, celebrated in Tampere in 2018. As its title suggests, it delves into the main theme of the conference, *corpus linguistics and the changing society*, thus bringing to the fore diachronic studies in which the interplay between linguistic changes and societal developments – be they cultural or technological – is explored. The volume comprises eleven original and thought-provoking chapters by both renowned and emerging scholars in the field of corpus linguistics. These contributions are organized into two separate parts, with the first dealing explicitly with linguistic changes that seem to respond to changes in the extralinguistic reality. Several chapters in this part of the volume also address the viability of corpora to examine the interrelation between language and society. The second part presents studies which survey language changes that are not directly connected to social advancements, but which are instead motivated by intralinguistic processes, including grammaticalization, among others.

Part I opens with **Martin Hilpert's** chapter, a call for caution regarding research on social change and its reflection in diachronic corpora. Hilpert first draws attention to five problems to bear in mind when connecting corpus findings to extralinguistic developments. These five problems are subsequently described, and compelling counterexamples are presented to show how corpus results might be misleading if these methodological pitfalls are not considered when designing our studies and analyzing our data. Then, a case study on the English *make*-causative construction is provided to implement the ideas previously discussed. Taking as a starting point Greenfield's (2013) claim that the diminishing power of interpersonal authority over time is reflected in the lexicon through a decrease in the frequency of words instantiating this concept (e.g., *authority*, *obedience*), Hilpert hypothesizes that the *make*-causative construction, which also expresses interpersonal authority (e.g., *don't make me marry him*), could mirror this change as well. However, after a systematic analysis that avoids the problems identified

\* Corresponding author: Daniela Pettersson-Traba, E-mail: [db.pettersson@ucm.es](mailto:db.pettersson@ucm.es)

before, it is shown that this particular function of *make* does not decline, but rather remains marginal over time. As a general conclusion, Hilpert encourages linguists to address the intersection between language and social change, but emphasizes the importance of doing so in a methodologically sound way to avoid drawing spurious conclusions. Therefore, Hilpert's fundamental contribution can serve to assess the validity of other studies that investigate social change through diachronic corpora.

**Gerold Schneider** and **Antoinette Renouf** also adopt a methodological focus in their respective chapters. Schneider investigates the changing views on poverty in the period 1500–1920. To this purpose, three different methods are employed and their validity to identify social change is assessed: the dictionary-based approach and two data-driven distributional semantic approaches (topic modelling and conceptual maps). A comparison of their results shows that the dictionary-based approach is not viable to draw conclusions about how attitudes towards poverty have developed diachronically. Rather, data-driven techniques must be used instead. In particular, according to the author topic modelling is more straightforward as it indeed serves to identify which topics poverty is associated with in different periods (e.g., religion, war, criminality). However, this method does not provide information about the association between individual words, only between words and general topics. Conceptual maps are specifically geared towards this purpose, but they are more intricate, and it is therefore difficult to interpret the results. Therefore, Schneider concludes that topic modelling is the most accessible method of the two.

Renouf also puts the validity of a method to the test, in this case different collocational analysis tools for the automatic identification of sense change. More specifically, she focuses on the emergence of new word senses in response to societal developments in areas including politics, technology, and gender issues. Five separate case studies are presented, each concentrating on an individual neoseme: *birther/birtherism*, *normalization*, *cougar*, *snowflake*, and *ghosting*. In each case study, the effectiveness of the tools is meticulously assessed and the author addresses the challenges posed by each neoseme, as well as the measures to tackle such challenges, which include expanding or reducing the context window and allowing or not for typographical variants, among others. Renouf argues that, despite data sparseness and the complex relations of some neosemes with previously existing senses, the tools are certainly effective, although the frequency of the node words needs to be high enough to draw valid conclusions.

The two remaining chapters of Part I, written by **Maura Ratia**, and **Gavin Brookes** and **David Wright**, revolve around the changing representation of certain groups of people and how this is manifested in language by recurring to collocational analysis. Ratia focuses on the depiction of patients in medical discourse throughout 1500–1800, a period which witnessed major developments in medicine and changes in moral values, moving towards a more polite and humane society. It is hypothesized that physicians' attitudes towards their patients might have been influenced by these developments in the sense of becoming more concerned about their well-being. To probe this hypothesis, Ratia examines the collocational behavior of *patient*. The paper provides a rigorous bibliographical review of the notion of collocation, and the choices made regarding criteria such as context window and frequency thresholds are well justified. Additionally, it is acknowledged that selecting other criteria would lead to different results (cf. Peirsman, Heylen and Geeraerts 2008 for similar arguments). Ratia demonstrates that in EModE patients are mainly viewed as objects whereas in LModE they are represented as experiencers. This change is noticeable in the occurrence in the later period of collocates with a negative connotation (e.g., *complains*, *aversion*), which point to a concern for the patients and their suffering. Another conclusion is that while in EModE the singular and plural forms behave similarly, in LModE the plural form increasingly appears with collocates related to hospital life (e.g., *hospital*, *admission*). This finding is claimed to be a result of the increasing significance of hospitals and public health in the 18<sup>th</sup> century.

Brookes and Wright turn their attention to the representation of non-native speakers of English living in Britain as reflected in the right-leaning press. In this case, a more recent and significantly shorter period is examined, namely 2005–2017. The authors examine the collocational behavior of the phrase *speak English* in different forms by combining corpus data and Critical Discourse Analysis. The extracted collocates are classified into four broad categories that represent the main themes of the

“speak English” debate: (i) proficiency, (ii) multilingualism, (iii) learning English and integration, and (iv) public services and the private sector. The results suggest that although non-native speakers of English are consistently portrayed in a negative light, the picture worsens as time progresses. This is reflected in the four categories of collocates distinguished. For instance, in the category of proficiency there is a shift in focus from specific individuals not being able to speak English well enough to the general immigrant population in later years. Therefore, the authors rightly reach the conclusion that there has been a widening in the 2010s in terms of both the scope and targeting of the stigmatization of immigrants by the right-leaning British press.

Three out of the six chapters in Part II explore a well-studied domain in linguistics, that of intensification, which has proven to be a phenomenon subject to constant change and renewal. **Karin Aijmer** and **Zeltia Blanco-Suárez** concentrate on the development of specific intensifiers. *Absolutely* is the object of study in Aijmer’s analysis, which traces its meaning and function in spoken PDE by drawing on data from BNC1994 and BNC2014. The findings show that the use of the intensifier varies across gender and age, but in a way that goes against assumptions in variationist sociolinguistics. This is so because it is not the younger generation who is behind the increase of *absolutely*, which is preferred by older speakers. Additionally, females are not the ones leading the development of *absolutely* from an intensifier with degree meaning (e.g., *absolutely brilliant*) to an emphaser of truth value (e.g., *absolutely love it*) and a discourse marker. The author concludes that the rise in frequency of *absolutely* goes hand in hand with syntactic, semantic, and pragmatic changes which point to a process of grammaticalization that is still ongoing.

Blanco-Suárez examines the long diachrony of the near-synonyms *deadly* and *mortal*, both of which can be used as boosters. The author traces their origin and development from OE to PDE by resorting to an extremely rich variety of resources and corpora. Interesting conclusions are drawn about the pathways of the two forms, demonstrating that they have both undergone processes of grammaticalization and (inter) subjectification, going from being used in a literal sense to an intensifying function, as in *deadly proud*. Nevertheless, differences in their developments are also identified: whereas *deadly* has a more marked negative prosody, *mortal* is more common with neutral and even positive collocates. In fact, although the adverbial forms of the two terms are more grammaticalized than their adjectival counterparts, this is especially true of *mortal*, which co-occurs relatively frequently with positive collocates in its intensifying uses.

A different approach to the study of intensifiers is adopted by **Martin Schweinberger**, who focuses on the distribution of amplifiers in American English from the 1850s to the 2000s. The most innovative aspects of this contribution are (i) the large number of amplifiers analyzed and (ii) the fact that a written genre is examined, namely fiction. A wide range of state-of-the-art statistical methods in corpus linguistics are used (e.g., linear regression, collexeme analysis). By examining the adjectives modified by the amplifiers, Schweinberger shows that changes occur only when they are used predicatively. In particular, *so* overtakes *very* from the 1980s onwards, which is surprising given that *really* has been claimed to be the competitor of *very* in spoken discourse in several varieties of English (e.g., D’Arcy 2015 on New Zealand English). However, this increase of *so* is not motivated by a broadening of its collocational profile, but due to its strong association with the high-frequency adjectives *sorry* and *beautiful*. Therefore, the author argues that this does not constitute a case of full lexical replacement of *very* by *so*.

**Yoko Iyeiry** sets out to examine the development of a class of adverbs which has to date not received much attention in the specialized literature: *-ingly* adverbs. The periods analyzed are Late Middle and Early Modern English, since most existing studies have focused on later periods. Although some findings are in line with previous research, such as the existence of a rise in frequency of *-ingly* adverbs in EModE, many of them contradict earlier claims about their development. First, given the amount of different types of *-ingly* adverbs in both ME and EModE, Iyeiry demonstrates that they are not as infrequent in these periods as previously assumed. In fact, adverbs such as *willingly* and *accordingly*, which are used recurrently by many authors, seem to be fairly well-established already in EModE. Second, Harry Potter adverbs, a class of adverbs of a rather elusive nature that tend to co-occur with “verbs of saying,

motion, and watching" (Broccia 2012: 151), do occur in Iyeiry's dataset already in ME and EModE, thus going against Broccia's (2012) conclusion that they are not attested until LModE. Therefore, the author convincingly argues that EModE appears to be a crucial moment in the development of *-ingly* adverbs, which highlights the importance of conducting more research on this period in future studies.

The two remaining chapters of the volume move away from the study of individual adverbs, either in isolation or in competition with others, to a focus on larger chunks. **Laurel J. Brinton** investigates the relatively little studied non-inference marker *that is not to say (that)*, which unlike its positive counterpart is not thoroughly discussed in reference grammars and dictionaries. Both synchronic and diachronic perspectives to the data are provided, and the use of the non-inference marker is examined in American and British English. Many sources are drawn on, given its low frequency until its rise in the 20<sup>th</sup> century, but there is still a gap in the data in the 17<sup>th</sup> and 18<sup>th</sup> centuries, in which *that is not to say (that)* is hardly ever attested. Brinton argues this gap to be due to its low frequency in fiction and the fact that most diachronic corpora of these two centuries consist mainly of fictional texts. As the author herself acknowledges, this constitutes a shortcoming of the study, making it difficult to draw conclusions about the diachronic development of the non-inference marker. Nevertheless, interesting findings emerge from the analysis. First, *that is not to say (that)* behaves rather differently from its positive counterpart regarding both function and form. Functionally speaking, it signals to the addressee that s/he should not infer some piece of information from the previous discourse. Formally speaking, *that is not to say (that)* exhibits more variants than *that is to say (that)*, occurring also with *this*, *which*, and *it* instead of *that* in initial position, in contracted form, and with or without the complementizer. Regarding the diachronic development of the non-inference marker, evidence points to it being a case of grammaticalization, fulfilling many of the criteria for this process (e.g., desemanticization, decategorization, (inter)subjectification). However, although it is increasingly used in contracted form and without *that*-complementizer, Brinton legitimately claims that it cannot be ruled out that these two developments reflect a process of colloquialization rather than grammaticalization.

Finally, **Turo Vartiainen** and **Mikko Höglund** revisit developments of transitivity from EModE onwards. They focus in particular on the transitive uses of the verb *sit* and test whether the Transitivity Hypothesis (Hopper and Thompson 1980), which has figured prominently in typological studies, can also be used to explain diachronic developments. In the first case study, the authors analyze the distribution of *sit* with the simple and self-reflexive constructions (*sit me down* vs. *sit myself down*), given that this verb persisted in time with simple reflexives as compared to other verbs. Findings point to the two reflexive strategies being used with *sit* until relatively late in the history of English. However, there did not seem to be much competition between the strategies as they were used in different contexts. In the second case study, a recent development of *sit down* is analyzed, namely its rise with non-reflexive transitive uses, as in *he sat me down*. The authors show that this construction has probably originated due to the overlapping use of *sit* with *set*. In the 17<sup>th</sup> century *set* started to be used with a new meaning, 'to cause or allow someone to descend from a vehicle', involving indirect causation. Given that *set* and *sit* were at this point often confused, it is hypothesized that the latter also became associated with this new sense, but over time it began to express direct causation with the meaning 'to physically assist someone to sit down'. In the 20<sup>th</sup> century, *sit down* experienced changes because it moved from being used only to express direct causation, although infrequently, to becoming a common expression of indirect causation, that is, simply guiding or inviting, rather than physically forcing, someone to sit down by means of gestures or speech. Vartiainen and Höglund's study therefore demonstrates that it is worth revisiting topics which have already been thoroughly addressed in the past by resorting to large diachronic corpora which have only emerged in the last decades. This is so because these resources can help to shed further light on more detailed information about a linguistic development, which can often go unnoticed in smaller datasets.

The volume is a welcome contribution to the field of corpus linguistics, especially to research dealing with historical data, which is indeed a flourishing and diverse one. In particular, the broad scope of the book is reflected in the diverse array of resources, approaches, and orientations adopted. First, it presents research based on a large number of different diachronic corpora (e.g., ARCHER, COHA,

EEBO, CLMET). Second, the volume covers papers that survey long-term changes taking place in previous periods of the history of English, as well as ongoing change in PDE. Additionally, given that recent changes are also addressed, not only written language is analyzed, but also spoken discourse. Third, from a methodological viewpoint, more traditional qualitative descriptive analyses appear alongside highly quantitative ones that explore innovative state-of-the-art techniques, which are becoming increasingly common in corpus linguistics nowadays. Lastly, the volume includes contributions that explore the interplay between language and society, but also studies in which language change is motivated by internal factors rather than external ones. On this note, a potential weakness of the volume lies precisely in its title, *Corpora and the changing society*, which can be slightly misleading, as it does not accurately reflect its contents. Some readers might find it surprising that only five out of eleven papers included actually deal with how social changes are reflected in corpora, while the rest deal solely with the level of language. Nevertheless, this might of course be because many linguistic changes are neither extralinguistically motivated nor do they go hand in hand with social developments. In fact, the inclusion of both types of changes can also be considered an advantage, as it makes the volume interesting to a wider readership by providing a broader perspective of language change. All in all, *Corpora and the changing society* will certainly be appealing to scholars interested in language change, both lexical and grammatical, and in many respects it also paves the way for future research at the crossroads of diachronic linguistics, corpus linguistics, and social change.

## References

- ARCHER = *A Representative Corpus of Historical English Registers* 3.2. 1990-1993/2002/2007/2010/2013/2016. Originally compiled under the supervision of Douglas Biber and Edward Finegan at Northern Arizona University and University of Southern California. Modified and expanded by subsequent members of a consortium of universities. Current member universities: Northern Arizona, Southern California, Freiburg, Helsinki, Uppsala, Michigan, Manchester, Lancaster, Bamberg, Zurich, Trier, Salford, and Santiago de Compostela. <https://www.alc.manchester.ac.uk/linguistics-and-english-language/research/projects/archer/>.
- Broccia, Cristiano. 2012. Oriented *-ingly* adjuncts in Late Modern English. In H. Sauer and G. Waxenberger (eds.), *English historical linguistics 2008. Vol. 2: Words, texts and genres* (Current Issues in Linguistic Theory 324), 147–164. Amsterdam: John Benjamins.
- CLMET = *The Corpus of Late Modern English Texts*. Compiled by Hendrik De Smet, Hans Jürgen Diller and Jukka Tyrkkö. <https://perswww.kuleuven.be/~u0044428/>.
- COHA = Davies, Mark. 2010–. *The Corpus of Historical American English: 400 Million Words, 1810–2009*. <https://www.english-corpora.org/coca/>.
- D’Arcy, Alexandra F. 2015. Stability, stasis and change. The *long durée* of intensification. *Diachronica* 32 (4): 449–493. <https://doi.org/10.1075/dia.32.4.01dar>.
- EEBO = *Early English Books Online*. <https://www.english-corpora.org/eebo/>.
- Greenfield, Patricia M. 2013. The changing psychology of cultures from 1800 through 2000. *Psychological Science* 24: 1722–1731. <https://doi.org/10.1177/0956797613479387>.
- Hopper, Paul J. and Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56 (2): 251–299. <https://doi.org/10.2307/413757>.
- Peirsman, Yves, Kris Heylen and Dirk Geeraerts. 2008. Size matters. Tight and loose context definitions in English word space models. In M. Baroni, S. Evert and A. Lenci (eds.), *Proceedings of the ESSLLI Workshop on Distributional Lexical Semantics: Bridging the gap between semantic theory and computational linguistics*, 34–41. Berlin: Springer.