REVIEWS

Ewa Jonsson and **Tove Larsson** (eds.), *Voices past and present – Studies of involved, speech-related and spoken texts. In honor of Merja Kytö*. Amsterdam: Benjamins, 2020. Pp. xiii + 348. ISBN 9789027207654.

Reviewed by Ivor Timmis, Leeds Beckett University

This edited volume, produced in honor of Merja Kytö, is divided into three parts, dealing with Early Modern, Late Modern and Present-day English. There are nineteen chapters in the volume, which, as Jonathan Culpeper puts it in the foreword, 'rightly celebrates her pioneering, inspiring achievements, the end of which we have certainly not seen' (p. xiii). Chapter 1 is by Ewa Jonsson and Tove Larsson, the editors. 'Voices of English: Tapping into records past and present' defines the territory of the book: the investigation of spoken, speech-related and interactive written texts in English from a synchronic and diachronic perspective. They speak of the need to counter the bias towards authoritative written texts in linguistics and of the benefits of following the road less travelled for language change and register studies. Accordingly, many of the chapters focus on non-canonical or even socially deprecated features of spoken language.

Part I. Early Modern English

In chapter 2, 'Pragmatic noise is Shakespeare's plays', Jonathan Culpeper and Samuel J. Oliver focus on pragmatic noise, which they define as 'semi-natural noises, such as *ah*, *oh*, and *ha*, that have evolved to express a range of pragmatic and discoursal functions' (p. 11). The main source they use is the *Enhanced Shakespearean Corpus*. The discussion of the genesis, annotation and tagging of this corpus is relevant to many researchers in historical linguistics. The authors focus on the social distribution of pragmatic noise, finding that female characters and characters from the middle of the social hierarchy use pragmatic noise most. The question Culpeper and Oliver pose is why pragmatic noise should be concentrated in these groups. Is it male stereotyping of females or the role of females as emotional commentators in the plays or both? This study is a fine example of how a canonical source can be used for the study of non-canonical features.

Dawn Archer and Alison Findlay also use a Shakespeare corpus in chapter 3, 'Keywords that characterise Shakespeare's (anti)heroes and villains', in this case the *Encyclopaedia of Shakespeare's Language Project*. This corpus is used to carry out a keyword analysis of seven Shakespearean characters in order to shed light on their feelings towards others and motivation to act. The results of the keyword analysis are

then compared with previous analyses of Shakespeare's portrayal of villainy. The chapter illustrates very clearly how a corpus technique can contribute to the analysis of literary characters (p. 45):

> keyword analysis can illuminate the linguistic patterns that give nuance to characters whose actions are morally reprehensible or questionable, at the same time as grounding previous (literary) understandings of these characters' thoughts, feelings and ambitions in (empirical) linguistic analysis.

We remain with Shakespeare for Juhani Rudanko's chapter 4 ('Revealing speech: Agentivity in Iago's and Othello's soliloquies'), which discusses agentivity in Iago's (six) and Othello's (three) soliloquies. The rationale for the focus on soliloquies, Rudanko argues, is that they are a window into the speaker's mind and view of the world at the same time. The definition of 'soliloquy' is problematised and there is a detailed discussion of both agency and agentivity. Agentivity, Rudanko (p. 59) argues, has both syntactic and semantic properties and 'makes it possible to provide a linguistic foundation for judgements that have sometimes been based on intuition alone'. As with the two previous chapters, Rudanko provides a very clear illustration of how corpus tools can be used to investigate literary questions in fruitful ways.

In chapter 5, 'Saying, crying, replying, and continuing: Speech reporting expressions in Early Modern English', Terry Walker and Peter J. Grund discuss speech reporting devices in Early Modern English as evidenced in prose fiction texts in the *Corpus of English Dialogues* (sections covering the 1560s to 1670s). Their specific interest is in variation over time in speech reporting verbs and in word order. Speech reporting devices, they note, are not neutral, 'they are varied in their semantic and structural makeup, and they are deployed strategically for a number of communicative purposes' (p. 64). From their analysis, they conclude (p. 65) that authors 'use speech representation in creative ways for purposes of plot development and characterization' and that the choice of device signals 'the speech reporter's interpretation and stance towards the reported speech and possibly the original speaker'. The authors make the important point that, while we cannot access authentic speech from the past, the way speech is represented is itself of significance and opens up avenues for research.

In 'Interjections in early popular literature: Stereotypes and innovation' (chapter 6), Irma Taavitsainen investigates the dual role of interjections (*alas, lo, O*) in early modern drama: to convey genuine feelings and to mark stereotypical responses to difficult situations. She analyses the discourse context of use of each interjection qualitatively, e.g. co-occurrence with comic turns of the plot and with other linguistic features such as deictic expressions. She concludes that interjections are a kind of pragmatic noise, which is genre-specific and that, when used in conjunction with proximal deictic expressions, interjections bring the events and characters close to the immediate experience of the readers. This chapter is another fine example of how revealing a focus on non-canonical features can be.

In chapter 7, 'Godly vocabulary in Early Modern English religious debate', Jeremy J. Smith takes us out of the world of literature, or, rather, into the world of political

literature, to discuss the language of two different communities of practice – one Protestant, the other Catholic, in the 1560s, as represented by three evangelist writers and three Roman Catholic writers. The focus is on 'ideologically-charged lexicons of theology and insult' (p. 95). Each community of practice, Smith notes (p. 95), made use of 'a coded English vocabulary, including words not usually seen as part of the semantic field of religion, to mark their distinctive discourse community'. In the case of prototypical evangelical texts, this led to a dual lexical strategy (p. 111):

> Prototypical evangelical texts, therefore deployed, along with (un)godly, such expressions as 'repentance' and 'sin', insulted enemies by calling them papists, and – very interestingly – also used less obviously theologically-marked lexemes such as 'comfortable', 'joy', and 'righteous'.

Smith draws an interesting and thought-provoking analogy between these sixteenth-century 'pamphlet wars' (p. 96) and the social media battles of the present day.

Matti Peikola contributes chapter 8, 'Patterns of reader involvement on sixteenth-century English title pages, with special reference to second-person pronouns'. In his study of sixteenth-century book titles, he also draws an analogy with present-day language, comparing the use of 'you' in book titles with the ubiquitous use of 'you' in modern advertising, both strategies aiming to involve the reader (or listener) in the text in question. The front matter of books, Peikola (p. 114) explains, 'developed into a promotional system' in the early modern period. Based on an analysis of the 'Digital data set of English titles 1501-1600', drawn from ProQuest's *Early English Books Online* (EEBO), Peikola (p. 118) notes that the pronoun 'you' '[c]an co-occur with explicit reference to the reader e.g. '(good) Christian reader' and with some kind of directive asking readers to confer or judge or haste to hear'. This reader involvement strategy typically co-occurs with imperatives and periphrastic future/modal expressions with *shall* and *may*. As with Smith's chapter, Peikola's analogy with modern-day texts is most thought provoking.

Part II. Late Modern English

Claudia Claridge uses the Late Modern English section of the *Old Bailey Corpus* to investigate the use of epistemic markers in chapter 9, 'Epistemic adverbs in the *Old Bailey Corpus*'. Epistemic adverbs, Claridge notes, 'mark the subjective assessment of a given speaker of their certainty relating to the matter at hand and the degree of commitment they want to communicate to interlocutors' (p. 134). The choice of epistemic marker by parties involved in a trial, e.g. j*udge*, *witness*, *accused*, *is*, *then*, clearly of great significance. Claridge studied the use of epistemic markers in the period 1720-1913, dividing this period into forty-year spans to facilitate diachronic analysis. There is not space to discuss the results in detail here, but Claridge reports that all the features studied increased in use over the whole period, leaving us with the intriguing observation (p. 151) that 'the speaker groups in the courtroom exhibit distinctive voices to a certain extent, which mirror in what way they are involved in the trial'.

Patricia Ronan also uses the *Old Bailey Corpus* in chapter 10, in this case to investigate question strategies in courtroom interaction in the early and late modern English period ('Question strategies in the *Old Bailey Corpus*'). The data was sampled at fifty-year intervals: 1732, 1780, 1830 and 1870. The author's interest is in the kinds of question used by different actors and in the degree of coercion a question strategy applies, e.g. a statement question, 'you murdered her' or a tag question, 'you murdered her, didn't you?' Ronan points out that in the early modern period defendants mostly represented themselves and used primarily yes/no questions, negative grammatical questions and *wh*-questions. The defendants, Ronan observes, remained active participants even in the late modern period. As far as lawyers are concerned, there was an increase in the variety of coercive questions in the nineteenth century as compared with the eighteenth century. As with the previous chapter, Ronan's chapter reveals how linguistic analysis can shed light on power relations in a specific social setting.

In chapter 11, '*Sure* in Irish English: The diachrony of a pragmatic marker', Raymond Hickey focuses on the use of *sure* as a pragmatic marker in Irish English, a use, he observes (p. 173), which has been '[l]ong recognised as indexical of Irish English before modern research'. The main data for the study comes from the drama texts in *A Corpus of Irish English*, with the *Corpus of English Dialogues, 1600-1700*, used for the purpose of comparison. As a pragmatic marker in Irish English, Hickey observes, *sure* functions 'as a pragmatic marker in sentence-initial or clause-initial position (occasionally in tag questions) expressing intersubjectivity in discourse' (p. 173). This use, Hickey points out, based on literary representations, dates back to the seventeenth century, and was at its most frequent in the nineteenth century. Pragmatically, Hickey argues, *sure* 'serves to reinforce shared knowledge and offers confirmation and/or assurance to one's interlocutor in discourse' (p. 175). Hickey leaves us with the fascinating speculation that the Irish use of the pragmatic marker *sure* might be related to a tendency in Irish rural conversation to seek consensus.

In 'American English *gotten*: Historical retention, change from below, or something else?' (chapter 12), Lieselotte Anderwald lays to rest the myth that *gotten* in American English is a historical retention of an earlier British form, describing it as 'a very curious case of an unintended by-product of prescriptivism' (p. 201). She argues rather that it was consciously promoted as a way to avoid *have got*, a form less favoured by 'careful writers' (p. 191). The criticism of stative *have got*, she argues, 'was so vicious that it may have affected all forms of *have got* and thus yielded to dynamic *have gotten*' (p. 198). Drawing on *COHA (Corpus of Historical American)*, Anderwald dates the revival of *gotten* to around the middle of the nineteenth century. The rise of *gotten* in American English in the nineteenth and twentieth centuries, she concludes, 'seems to illustrate the unusual case of change from above' (p. 201).

Part III. Present-day English

David Denison brings us into Present-day English in Chapter 13 ('Explaining explanatory *so*') with a very modern phenomenon: the use of *so* as 'a discourse particle in turn-initial

position which has such functions as accepting an invitation to take the floor and prefacing an explanation' (p. 207). Denison refers to this use as 'explanatory *so*'. The diachronic analysis is based on the spoken component of the 1994 and 2014 components of the *British National Corpus* (BNC). There is also reference to BASE (*British Academic Spoken English*) and a brief look at broadcasting data. Denison shows that, based on the BNC, the turn-initial use of *so* has approximately doubled in frequency from 1994 to 2014 and that its main function is to introduce an invited explanation. It appears, however, that it is almost absent from BASE in this function. Denison concludes that '[t]urn-initial *so* has been on a long-term trajectory of increase at the expense of *well* and other discourse particles in speech' (p. 220). It has provoked some annoyance, Denison suggests, as the 'conservative user does not themself have explanatory *so* in their repertoire, but it is highly salient to them when heard, coming very typically as the very first word' (p. 221)'

Chapter 14, 'Return to the future: Exploring spoken language in the BNC and BNC2014', by Ylva Berglund Prytz, reports on a diachronic study of future expression based on the 1994 and 2014 spoken components of the BNC, focusing on the 'frequency and proportions of the different expressions and collocations with personal pronouns' (p .227). The corpora were analysed using CQPweb (Lancaster University). Intriguingly, the results shows that future expressions are less frequent overall in 2014 than they were in 1994. Prytz concedes, however, that, while this may mean that ways of expressing the future have changed, it could also be the result of the different types of text sampled in each corpus. Prytz also shows that the BNC 1994 has a higher frequency of future-referring terms than the 2014 version. As is the case with so many chapters in this book, the author leaves us with food for thought: 'is it the case that when we choose to [refer to the future], we use other constructions than those used by speakers in 1994, foregoing also the future-referring terms here?' (p. 234).

Karin Aijmer, the author of chapter 15 ('*Sort of* and *kind of* from an English-Swedish perspective') investigates the hypothesis that the type nouns *kind of* and *sort of* 'have a meaning potential rather than a fixed meaning' (p. 248). The question is addressed from an English–Swedish perspective using the fiction texts from a parallel corpus and *Translation Corpus Explorer* for analysis: Aijmer argues that '[a] contrastive perspective is interesting if languages have markers which seem to have followed the same development' (p. 247). From the analysis, Aijmer concludes that '[a]s can be expected from their association with spoken involvement and interaction, *sort of* and *kind of* do not have a fixed meaning but their meanings depend on the context, in particular the syntactic co-text' (p.261). Aijmer identifies approximation and categorizing as core functions of these type nouns along with sub-functions of imprecision, downtoning and hedging, and observes that '[t]he development from taxonomic nouns to approximators can be observed in several Romance and Germanic languages' (p. 248). This chapter is a good example of how a finely focused study can open up wider reflection, which is a feature of this book.

Chapter 16 is by Anna-Brita Stenström, 'From *yes* to *innit*: Origin, development and general characteristics of pragmatic markers'. She investigates the use of the pragmatic

markers *yes*, *yeah*, *okay* and *innit*, with a particular focus on *innit* and how far it matches the functions of the other three pragmatic markers. There is a diachronic element to the research as the BNC 1994 and the *Bergen Corpus of London Teenage English* (COLT) are used as well as the more recent *Multicultural London English Corpus* (MLE) and BNC 2014. The BNC data includes the conversation of adults. Stenström discusses the pragmatic functions of each marker and how far they differ and overlap, arguing that '[s]everal factors contribute to the functions of *yes, yeah, okay* and *innit* in the discourse, such as who is talking to whom, what the conversation is about, in what context the pragmatic marker occurs, and not least its position in the utterance' (p. 269). The diachronic picture differs according to whether we are dealing with youth or adult use (p. 280): 'the winners are *yeah*, *okay* and *innit* in the youth corpora versus *yeah* and *okay* in the adult corpora'. This chapter is a good example of how a detailed study of a socially deprecated form can reveal system in its use.

Sarah Schwarz and Erik Smitterberg also focus on features that might be regarded as non-canonical in chapter 17, entitled '"If anyone would have told me, I would have not believed it": Using corpora to question assumptions about spoken vs. written grammar in EFL grammars and other normative works'. These features are split infinitives, *like* (for 'as if'), *though* as a conjunct and *would have* in conditional clauses. They used *COCA 2019* for their research and the target feature was searched in the spoken component and in at least one written genre. The 'headline' findings are as follows: (i) split infinitives are almost as common in academic writing as they are in speech, (ii) the 'as if' function of *like* is well established in academic writing, (iii) *though* as a conjunct is quite well established in academic texts, and (iv) *would have* in conditional clauses is mainly confined to speech. Schwarz and Smitterberg make two important points for teachers. Firstly, we need to question both our intuition and the grammatical 'rules' in the books we use as regards what is acceptable in written English. This struck a chord with me as I remember confidently telling a student in the 1990s that you could not use *access* as a verb. Secondly (p. 297) 'there is a difference between proscribing certain features and making learners aware of perceived differences and empowering them to make their own choices'.

In 'Intensification in dialogue vs. narrative in a corpus of present-day English fiction' (chapter 18), Signe Oksefjell Ebeling and Hilde Hasselgård compare adverbial intensification of adjectives in the narrative and the dialogue sections of fiction. The data they used is from the original fiction part of *the English–Norwegian Parallel Corpus*. They found evaluative adjectives to be more frequent in dialogue than narrative while narrative uses a wider range of such adjectives. They also considered the type and position of the intensifier, finding that 'amplifiers and predicative function are the preferred choice in both dialogue and narrative, and … these are comparatively more common in dialogue' (p. 307). More generally, Ebeling and Hasselgård report that amplifiers are twice as frequent in spoken English as compared with written English and that the lexical choice of amplifier differs in spoken and written English. The chapter is an interesting example of a study in inter-register variation.

Finally, chapter 19 'Orality on the searchable web: A comparison of involved web registers and face-to-face conversation', by Douglas Biber and Jesse Egbert, address a question which is straightforward to pose but far from straightforward to answer: how

far do web registers represent the linguistic characteristics of spoken registers? They note that the relationship between colloquial written registers and speech has long been of interest (it was certainly of relevance to my own study, of nineteenth-century pauper letters (Timmis 2020)). The precise question Biber and Egbert investigate is '[are there] written registers on the searchable web, readily accessible to researchers and practitioners, which provide reasonable representations of the typical discourse style found in spoken interaction?' (p. 319). Their corpus consists of 44,000 documents extracted from the *Corpus of Web-based English* (GloWbE). The analysis they carry out is highly register-sensitive and leads to the conclusion that there is a continuum of orality, but no single register is a close match for speech, as the linguistic characteristics of register are crucially shaped by situational factors (p. 331):

> … although song lyrics, transcribed interviews, TV transcripts, and discussion forums are the most 'oral' of the registers found on the public searchable web, their situational characteristics differ in several key respects from both spoken conversation and (super) synchronous CMC. It turns out that that these situation differences correspond to systematic linguistic differences.

The hallmark of all these chapters is rigorous methodology and a bold originality which combine to encourage the reader to look at language in new ways and from different perspectives, a fitting tribute indeed to the work of Merja Kytö.

Reviewer's address:
2 Victoria Crescent
Leeds LS18 4PR
UK
i.timmis@leedsbeckett.ac.uk

Reference

Timmis, Ivor S. 2020. *The discourse of desperation: Late nineteenth and early twentieth century letters by paupers, prisoners and rogues*. London: Routledge.

(Received 11 October 2021)

Sofia Rüdiger and **Daria Dayter** (eds.), *Corpus approaches to social media* (Studies in Corpus Linguistics 98). Amsterdam and Philadelphia: John Benjamins, 2020. Pp. vi + 210. ISBN 9789027207944.

Reviewed by Mikko Laitinen ⓘ, University of Eastern Finland

Corpus approaches to social media contains a collection of chapters based on the papers presented in a workshop at the 40th ICAME conference (International Computer Archive