

Review of Elena Seoane and Douglas Biber eds. 2021. *Corpus-based Approaches to Register Variation*. Amsterdam: John Benjamins. ISBN: 978-9-027-21054-8.
<http://doi.org/10.1075/scl.103>

Claudia Claridge
University of Augsburg / Germany

This book is a very rich collection, approaching register from various angles and with a wealth of methodologies and tools. Data types and contexts focused on in the contributions include modern (especially Chapters 5 and 6) and historical varieties of English (Chapters 10–12), as well as the analysis of English as a Foreign Language (EFL; Chapters 8–9). Both multiple- and single-register research (e.g., science, pop lyrics, or newspaper writing) is included, as well as studies approaching register from the point of view of one particular phenomenon (e.g., dative alternations, noun phrase modification). Needless to say, multi-dimensional analyses (MDA) are found here, both in the original (Chapter 9) and in the additive sense (cf. Berber Sardinha 2014, Chapter 8). A host of other methods are employed in these studies, such as mixed effects regression models (Chapter 3), random forests (Chapter 5), geometric multivariate analysis (Chapter 6), or generalised linear models (Chapter 9), just to name a few. Some tools receive a surprising but useful employment, such as the orthographic regulariser VARD (Baron and Rayson 2008) developed for historical data, being used here on modern pop lyrics (Chapter 8).

The majority of contributions are based on the understanding of the term ‘register’ by Biber (1988) and Biber and Conrad (2019), namely as “varieties associated with particular situational contexts that can be characterised for their audiences, medium [...], interactivity, production circumstances, communicative purposes” (p. 2), and thus

with similar linguistic characteristics. While the latter have been investigated in quantitative ways, the characteristics of the situational contexts have neither been measured nor treated as continuous variables but have usually been described in a subjective and generalising manner. A critical discussion of such earlier approaches together with a comparison of the textlinguistic (Biber 1988 and following) and a systemic-functional perspective is presented by Biber, Egbert, Keller, and Wizner in Chapter 2, before they proceed with their undertaking of providing such measurements of the situational context. They report on two case studies from earlier papers (Biber *et al.* 2020 and Biber *et al.* 2021): one based on a range of web texts and one on conversational discourse types in the spoken component of the second *British National Corpus* (Spoken BNC2014; cf. Love *et al.* 2017). The web case study quantified 23 situational parameters on a six-point scale averaged across two independent raters and resulted in two dimensions of situational contexts. While Dimension 1 is characterised as opinionated discourse vs. technical information supported with evidence, Dimension 2 marks narrative, entertaining discourse versus other communicative purposes (explanatory, advice, and procedural discourse), from which dimension scores for individual texts can be derived. The scored texts form five clusters, which actually cut across registers. The second case study coded conversational discourse units for nine communicative purposes, leading to 16 clusters, characterised by labels such as ‘figuring-things-out’, ‘joking around’, or ‘conflict’. While an overall convincing start at the situational aspect, the second case study would have profited from more detail and illustration.

Two contributions (Chapters 3 and 5) deal with the dative alternation, which might have been placed in direct sequence. Chapter 3, by Engel, Grafmiller, Rosseel, Szmrecsanyi, and van de Velde, investigates and compares the effects of register and of various language-internal constraints on the choice of the dative realisation as either ditransitive or prepositional. Taking into account seven language-internal factors, in particular, recipient and theme definiteness as well as constituent length, and four registers as predictors (conversation, parliamentary debates, blogs, and newspaper articles) in a mixed effects regression analysis, they showed core grammar to be relatively stable across registers and register effects to be smaller than for other factors. Similarly, in Chapter 5 by Röthlisberger, register is marginally outranked in importance by variety of English (of which nine are investigated) and in some varieties (e.g., British

and Singapore English) there is little inter-register variation. Register differences across varieties are generally small, but with subtle distinctions regarding formal registers, which is hypothesised to be due to indigenisation effects. The internal constraints weight ratio and pronominal recipient turn out to be the most important factors overall. While the double object construction generally dominates, the prepositional variant is more likely in non-native Englishes.

Chapters 4 and 6 are two outliers in the volume in the sense that they both proceed from a Hallidayan systemic-functional perspective on register. In Chapter 4, Pérez-Guerra tests the hypothesis of theme choice being indicative of register, on the basis of themes having dual linguistic and situational/functional status just like registers. Secondly, the adequacy of two definitions of theme, those by Halliday (1985) and by Berry (1995) respectively, is tested as to their contribution to register characterisation. In an analysis encompassing 15 written registers of American English, the Hallidayan concept of theme (first ideational element) is claimed to be a plausible predictor of (dis)similarity between registers, while Berry's preverbal theme concept fares less well. The chapter would have gained in persuasiveness if the technical description had been somewhat more accessible and, in particular, if more linguistic illustration had been provided.

Neumann and Evert (Chapter 6) used 41 register-sensitive lexico-grammatical features for their analysis of 2,844 texts from the Hong Kong, Jamaica, and New Zealand components of the *International Corpus of English* (ICE).¹ They use a geometric multivariate analysis, inspired by multidimensional analysis (MDA), to explore and visualise the linguistic differences between texts. The resulting four dimensions are: 1) conceptual speaking/conceptual writing (e.g., ICE categories conversation, social letters, and news), 2) dialogic written/neutral (e.g., social letters, creative writing), 3) descriptive-narrative versus instructive-regulative (e.g., news, business letters, and administrative writing), and 4) neutral/online production (e.g., unscripted discourse). The three varieties differ in variance across the four dimensions, with the least variance in New Zealand and the most variance in Hong Kong texts, which may be due to less established conventions in the younger variety. Tenor-related (pragmatic) aspects seem to contribute more to variation than field-related aspects. The study is interesting not only for the visualisation aspect, but also for the overlap and

¹ <http://ice-corpora.net/ice/index.html>

differences it shows regarding Biber's dimensions, and for the potential problems inherent in the ICE text classification that it highlights. This very rich treatment also makes repeated reference to web materials for further illustration and corroboration, which, on the one hand, is laudable, but, on the other hand, detracts from the independence of the chapter.

Chapter 7, by Botha and van Zyl, focuses on the noun phrase, a feature that has previously been attested with interesting behaviour in variation and change. The novel approaches in this contribution concern using proportions of modifiers relative to the number of nouns used in a register and conducting as many as 45 pairwise comparisons for individual modifier forms in ten registers using effect size measures. The data combines the five registers of the *Corpus of Contemporary American English* (COCA; Davies 2008–) with five web-based registers from the *Corpus of Online Registers* (CORE; Davies 2016), which are matched for similar communicative purposes and intended audiences. Modification structures included in the study are premodifying (proper) nouns and adjectives (L1 position only) and postmodifying prepositional phrases, non-finite and relative clauses. High levels of premodification as well as of prepositional and non-finite postmodification were found to characterise written informational registers, while postmodification by *that*-relatives marked oral and involved registers. Some modification features showed more register-sensitivity than others, for instance, pronominal (proper) nouns and *which*-relatives as opposed to adjectives and *that*-relatives. The high number of comparisons allowed the observation of very fine-grained differences.

The next two chapters have a common focus in so far as both are concerned with EFL matters. Werner's contribution on pop lyrics points to them being a register in their own right, in contrast to prevailing views (especially EFL) that they are speech-like and conversational in nature. An 'additive MDA' (Berber Sardinha 2014) performed on the *Corpus of English Pop Lyrics* (LYPOP; Werner 2020) from 2001 to 2016, comprising 1,842 lyrics and 547,758 tokens, mapped the register onto Biber's (1988) original dimensions. In line with their non-conversational situational characteristics, pop lyrics usually have closest score associations with written (all dimensions) and partly with formal and informational types, such as official documents (Dimension 2) or academic prose (Dimension 4). Nevertheless, the large standard deviations exhibited by pop lyrics scores always include (Dimensions 3, 4, and 5) or overlap substantially with

conversation (Dimensions 1, 2, 6). Werner concludes that while the pop lyric register is not conversational as such it uses a range of features in such a way as to produce the impression of an imagined speech event with pseudo-dialogicity and thus a pretence at conversationality.

Proceeding from the fact that little is known about EFL academic learner writing beyond performance in the register of argumentative essays, the chapter by Larsson, Paquot, and Biber reports a new MDA to investigate register effects in EFL learner writing. The aim is to find out how it differs both from native writing and across different L1-groups and also to investigate how these findings are influenced by register. A MDA performed on a 3.5 million word corpus of native and EFL argumentative essays, research papers, and published scientific articles (all drawn from existing corpora such as the *International Corpus of Learner English* [ICLE], the *Louvain Corpus of Native English Essays* [LOCNESS] the *Varieties of English for Specific Purposes dAtabase* [VESPA], the *Louvain Corpus of Research Articles* [LOCRA], among others)² led to two dimensions, Dimension 1 distinguishing a personal versus a topic-focused style and Dimension 2 an evaluative style as opposed to factual description. On both dimensions the influence of register is shown to be more important than either (non-)nativeness or the specific L1 of learners. A more personal style thought to be a generic characteristic of EFL writing is only found for argumentative essays (including those of native speakers), while all writers show register awareness by adopting a more topic-focused (Dimension 1) and factual-descriptive (Dimension 2) style for research papers. Moreover, EFL learners are not a coherent group, but show significant differences between different L1 backgrounds: for example, while Norwegians prefer a more personal approach, French learners use more topic-focussed writing.

The final three papers in the collection all take a diachronic perspective. Rodríguez-Puente's chapter charts the attestations of nominalisations with nine Romance and native suffixes covering four meanings across 18 registers of Early Modern English (1500–1760), taken from the *Corpus of English Dialogues* (CED; Kytö and Walker 2006), the *Penn-Helsinki Parsed Corpus of Early Modern English* (PPCEME; Kroch *et al.* 2004), and the *Corpus of Historical English Law Reports*

² <https://uclouvain.be/en/research-institutes/ilc/cecl/corpora.html>

(Rodríguez-Puente *et al.* 2018). The suffixes include highly frequent *-ion*, medium-frequent *-ment*, *-ity*, *-ness*, *-age*, *-ship*, and low-frequent *-dom*, *-hood*, and *-head*. Their occurrence is significantly linked to register, with formal, writing-based and writing-purposed texts showing higher and informal, speech-related texts lower frequencies. Exceptions to this pattern are due to the type and purpose of text, with a narrative style leading to lower frequencies (e.g., the bible, fiction, or travelogue) and an instructional and persuasive outlook to higher frequencies (e.g., sermons). While nominalisations increase in most registers, the most in sermons (1640–1710), in line with the increasing nouny-ness and literate character of texts in the period, private letters and trial proceedings show a decrease. The developments are shown nicely in Figure 4, which, however, seems to be lacking the last period for the registers drama, trial proceedings, and witness depositions. The productivity of suffixes, shown by aggregation of types from the first to the last period, is most pronounced for borrowed suffixes overall, but also for *-ness*. There are generally no register effects regarding productivity, with trials again standing out and showing an unusual decline.

Degaetano-Ortlieb's contribution is an investigation into the development of the scientific register in the twentieth century, here represented by the mathematical, physical, and engineering publications of the *Proceedings of the Royal Society of London*.³ In a bottom-up and data-driven approach based on POStrigrams and using Kullback-Leiber divergence, critical periods of change as well types of change were identified. The (early) 1920s turned out as an important period of change, followed by later stabilisation, while the crucial constructions in the registerial change all involve nominal compounds (in particular, det-N-N, N-N-prep and adj-N-N). With premodified noun phrases rather than those with prepositional postmodification becoming a more distinctive use, the change also represents informational densification. Also, the rise of pure N-structures instead of N+prep structures shows not only higher informativity but also increasing specialisation.

The final contribution, by Hiltunen, deals with sub-register variation in nineteenth-century newspapers, but also with the question of how to work best with the *British Library Newspapers*⁴ database in a corpus-linguistic approach. Regarding the latter, Hiltunen extracted two corpora (A and B) automatically from the database for the

³ <https://royalsocietypublishing.org/journal/rspl>

⁴ <https://www.bl.uk/collection-guides/newspapers>

purpose of data triangulation to overcome weaknesses of individual corpora. Corpus A (4.9 million words) sampled whole issues from five geographically diverse papers published during the whole century by choosing issues from two months in 10-year intervals. Corpus B (10.1 million words) was compiled from 100 texts for each of the seven sub-registers in focus extracted for every decade. Both corpora also required an Optical Character Recognition (OCR) confidence level of 90 per cent for the texts to be included. It was especially this essential criterion for linguistic accuracy which led to massive loss of data (e.g., one whole newspaper for Corpus A) and caused Corpus A to be very unbalanced. As for Corpus B, even though it was more balanced, it also showed some coverage gaps regarding sub-genres (e.g., sports, classified ads). The POS-tagged corpora were used to carry out a synchronic register analysis with selected features chosen from Biber's (1988) Dimensions 1, 2, 4, and 5, namely, private verbs, first, second and third person pronouns, past tense forms, suasive verbs, infinitives, conjuncts, and sequences of two proper nouns. All of those indicated sub-register differences, with most marked distinctions in editorials for involved, persuasive, and explicit-linkage features, and with news and sports characterised by narrative features. As an outlier, birth/death/marriages notices were only characterised by proper noun sequences.

The entire volume clearly makes for very stimulating reading, with many convincing insights and with inspirations for data sources and use, as well as for methodology. Sometimes, however, the technical details regarding the latter become somewhat overwhelming and not easy to follow for the reader —a minor criticism that applies more to papers in the first half of the volume. Another minor weakness of the volume concerns the general lack of cross-references, for which there would have been ample opportunity given the very real links between the contributions. One striking example of a missing cross-reference concerns Matthiessen's (2019: 26) map of register traditions mentioned in Chapter 6 (p. 145), which is actually reproduced in Chapter 4 in the volume (p. 88).

REFERENCES

- Baron, Alistair and Paul Rayson. 2008. VARD 2: A tool for dealing with spelling variation in historical corpora. In *Proceedings of the Postgraduate Conference in Corpus Linguistics*. Birmingham: Aston University. <http://ucrel.lancs.ac.uk/people/paul/publications/BaronRaysonAston2008.pdf>
- Berber Sardinha, Tony. 2014. 25 years later: Comparing internet and pre-internet registers. In Tony Berber Sardinha and Marcia Veirano Pinto eds. *Multi-dimensional Analysis, 25 Years on: A Tribute to Douglas Biber*. Amsterdam: John Benjamins, 81–105.
- Berry, Margaret. 1995. Thematic options and success in writing. In Mohsen Ghadesy ed. *Thematic Development in English Texts*. London: Pinter, 55–84.
- Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, Douglas and Susan Conrad. 2019. *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Biber, Douglas, Jesse Egbert and Daniel Keller. 2020. Reconceptualizing register in a continuous situational space. *Corpus Linguistics and Linguistic Theory* 16/3: 581–616.
- Biber, Douglas, Jesse Egbert, Daniel Keller and Stacey Wizner. 2021. Towards a taxonomy of conversational discourse types: An empirical corpus-based analysis. *Journal of Pragmatics* 171: 20–35.
- Davies, Mark. (2008–). *The Corpus of Contemporary American English (COCA): 560 Million Words, 1990-Present*. <https://corpus.byu.edu/coca/>
- Davies, Mark. (2016–). *Corpus of Online Registers of English (CORE)*. <https://www.english-corpora.org/core/>
- Halliday, M. A. K. 1985. *An Introduction to Functional Grammar*. London: Edward Arnold.
- Kroch, Anthony, Beatrice Santorini and Lauren Delfs. 2004. *The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME)*. Department of Linguistics: University of Pennsylvania.
- Kytö, Merja and Terry Walker. 2006. *Guide to a Corpus of English Dialogues 1560–1760*. Uppsala: Acta Universitatis Upsaliensis.
- Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina and Tony McEnery. 2017. The Spoken BNC2014: Designing and building a corpus of everyday conversations. *International Journal of Corpus Linguistics* 22/3: 319–344.
- Matthiessen, Christian M.I.M. 2019. Register in systemic functional linguistics. *Register Studies* 1/1: 10–41.
- Rodríguez-Puente, Paula, Teresa Fanego, María José López-Couso, Belén Méndez-Naya, Paloma Núñez-Pertejo, Cristina Blanco-García and Iván Tamaredo. 2018. *The Corpus of Historical English Law Reports 1535–1999 (CHELAR)*, v.2. University of Santiago de Compostela: Research Unit for Variation, Linguistic Change and Grammaticalization.
- Werner, Valentin. 2020. Teaching grammar through pop culture. In Valentin Werner and Friederike eds. *Pop Culture in Language Education: Theory, Research, Practice*. London: Routledge, 85–104.

Reviewed by
Claudia Claridge
University of Augsburg
Faculty of Philology and History
Universitätsstraße 10
86159 Augsburg
Germany
E-mail: claudia.claridge@philhist.uni-augsburg.de