

### References

- Aarts, Bas. 2004. Grammatici certant. Review Article of Rodney Huddleston and Geoffrey Pullum (2002) *The Cambridge grammar of the English language*. *Journal of Linguistics* 40(2), 1–18.
- Huddleston, Rodney. 1984. *Introduction to the grammar of English*. Cambridge: Cambridge University Press.
- Huddleston, Rodney & Geoffrey Pullum *et al.* 2002. *Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Nikolaeva, Irina. 2007. Introduction. In Irina Nikolaeva (ed.), *Finiteness: Theoretical and empirical foundations*, 1–19. Oxford: Oxford University Press.
- Palmer, Frank. 1965/1987. *The English verb*. London: Longman.

(Received 11 November 2022)

doi:[10.1017/S1360674322000478](https://doi.org/10.1017/S1360674322000478)

**Elena Seoane and Douglas Biber (eds.)**, *Corpus-based approaches to register variation* (Studies in Corpus Linguistics 103). Amsterdam and Philadelphia: John Benjamins, 2021. Pp. xi + 341. ISBN 97827210548.

Reviewed by Marcia Veirano Pinto , Universidade Federal de São Paulo

Register, understood as culturally recognizable categories defined by their production circumstances, finds itself in the foreground of corpus studies today. One of the reasons for such prominence is that the body of research on register has recently revealed that many, if not most registers, are not broad, neat, closed categories. However, the blurriness of their boundaries does not render registers immaterial, quite the contrary, as their production circumstances tend to be stable (p. 57). What such a characteristic does is to give them a continuous makeup that helps speakers narrow down their linguistic choices. This role of registers is widely accepted within the academic community and makes register variation studies, in the words of Elena Seoane and Douglas Biber, ‘a vast domain’ (p. 4). The different theoretical frameworks used in the chapters of *Corpus-based Approaches to Register Variation* (henceforth *CARV*) – namely Probabilistic Grammar, Systemic-Functional Linguistics and Information Theory – present ways corpus linguistics can greatly contribute to our understanding of the role of registers in language, either as an approach or a method. Thus, the book is not for the reader who believes that corpus-based studies are homogeneous and follow a rigid framework. It is for the reader who is interested in learning what we know so far about the relationship between registers and the linguistic choices we make, when we both communicate with others and negotiate relationships. It is also for the reader who questions the notion of registers as clear-cut universal categories that can be defined *a priori* and would like to have a better understanding of the implications of such a view.

On reading *CARV*, what becomes evident is that there is little point in describing language patterns without taking into consideration the registers where they occur, as such description will not help us understand the reasons behind speakers' linguistic choices. This point is raised and discussed in every chapter in *CARV*. The first chapter in the volume, written by the editors Seoane and Biber, has the same title as the book: 'Corpus-based approaches to register variation' (pp. 1–18). It provides a comprehensive historical background for register variation studies and for empirical register research, ending with notes on the different theoretical and methodological frameworks as well as the corpora used by the scholars who contribute to the volume.

Chapter 2, 'Extending text-linguistic studies of register variation to a continuous situational space: Case studies from the web and natural conversation' (pp. 19–50), by Douglas Biber, Jesse Egbert, Daniel Keller and Stacey Wizner, follows the Text-Linguistic approach. The authors discuss the theoretical grounds of the situational analysis of registers, suggesting that the methods used so far would profit from an empirical investigation of the variation in the situational characteristics of the texts that belong to a specific register. To that end, they introduce methods for treating and coding situational variables as a continuum, to identify dimensions of situational variation. This new methodological framework is applied to two samples of texts: web texts taken from the *Corpus of Online Registers of English* (CORE) and face-to-face conversation texts taken from the spoken portion of the *British National Corpus 2014* (Spoken BNC2014). As a result, they identify five situationally well-delimited web text types and nine higher-level face-to-face conversation discourse types, which are described according to their communicative purposes. The authors justifiably say that the greatest strength of this situational analytical framework is that it allows for the description of texts on the web 'that do not belong to any culturally-recognized category' (p. 45).

The third chapter in *CARV*, 'How register-specific is probabilistic grammatical knowledge? A programmatic sketch and a case study on the dative alternation with *give*' (pp. 51–84), by Alexandra Engel, Jason Grafmiller, Laura Rosseel, Benedikt Szmrecsanyi and Freek Van de Velde, adopts a Probabilistic Grammar framework. The authors' aim is to advocate for the inclusion of register in variationist linguistics, by proving its role as a modulator of language-internal constraints on the grammatical choices made by speakers. To achieve such an objective, they combine a corpus study of extensively annotated texts representing four registers – namely spoken informal (Spoken BNC2014), spoken formal (*House of Common Corpus 2007–2014*), written informal (the British English blog portion of GloWbE) and written formal newspaper articles (from the newspaper *The Independent 2016–2019*) – with rating task experiments. The corpus study consists of applying logistic regression to utterances and sentences containing the verb *give*, extracted semi-automatically and coded manually for context. Various probabilistic effects are found for the formal and informal spoken registers, but not for the comparison between spoken and written registers. To the authors, this outcome may have resulted from the heterogeneity in the written texts in the corpus, as they have come from blogs and newspaper articles with multiple communicative purposes. Nonetheless, in view of the results for spoken registers, they conclude that

including register and experimental data in variationist linguistics is warranted and may help researchers to better understand the cognitive system related to probabilistic choice-making.

In chapter 4, 'Theme as a proxy for register categorization' (pp. 85–110), Javier Pérez-Guerra uses the Systemic Functional Linguistics framework to look at the premise that the thematic design of clauses – as defined by both Halliday & Matthiessen (2014) and Berry (1995) – can be used as a predictor of register membership. The corpus used in the study is composed of 90,258 clauses extracted from the *Crown* corpus with the *Stanford Parser*. Through a hierarchical cluster analysis carried out with the linguistic characteristics present in the themes of these clauses, Pérez-Guerra concludes that Halliday & Matthiessen's (2014) definition of theme – 'first (ideational) element' (p. 85) – is a better methodological and interpretational dissimilarity metric. The author also claims that the situational differences between registers, which include differences in participants and communicative purposes, promote the use of different linguistic features and strategies. To Pérez-Guerra, this claim finds support in the clustering of popular and speech-like registers against that of learned, formal ones.

Melanie Röthlisberger's chapter, 'Between context and community: Regional variation in register effects in the English dative alternation' (pp. 111–42), uses a Probabilistic Grammar framework to investigate the influence of both register and the regional background of speakers on dative alternation. To this end, she builds a corpus with samples from the *International Corpus of English* (ICE) containing nine national varieties: British, Canadian, Irish, New Zealand, Hong Kong, Indian, Philippine, Singapore and Jamaican English. The samples for each variety include 500 texts of 2,000 words each, representing spoken (40%) and written language (60%). To complement her data, she adds a sample of texts – mirroring the structure of the sample taken from the ICE corpus and totalling 500,000 words – extracted from the GloWbE corpus. She divides the texts in her data into four registers, namely spoken vs written and formal vs informal. The 13,171 variable dative tokens extracted from this corpus were analysed first by means of a conditional random forest and then by a mixed-effects logistic regression, to 'assess the effect of register on dative choice and the inner coherence of registers across the nine national varieties' (p. 121). Results from the conditional random forest indicate that national variety is a better predictor for the choice of dative variant than register. The regression model suggests that region-specific situational contexts are more likely to account for speakers' choices. In terms of registers, formality is a better predictor than mode, the degree of inter-register variation is more salient in some national varieties than in others, and linguistic choices in a specific register are marginally related to regional variation.

The sixth chapter, 'A register variation perspective on varieties of English' (pp. 143–78), by Stella Neumann and Stefan Evert, adopts a multidimensional approach to identify dimensions of register variation across national varieties of English. However, instead of using multidimensional analysis, they apply geometric multivariate analysis to a corpus composed of 2,844 texts, distributed among twenty registers, sampled from the Hong Kong, Jamaica and New Zealand English varieties in the ICE corpus. Their analysis is based on forty-one lexicogrammatical features pertaining to systemic

functional register theory, which are summarized in Matthiessen (2019). Following such methodology, Neumann & Evert identified four dimensions of variation, namely Conceptual Speaking vs Conceptual Writing, Dialogic Written vs Neutral, Descriptive-narrative vs Instructive-regulative and Neutral vs Online Production. Results show differences among the varieties of English and the modes (speech and writing) in each of the four dimensions as well as continuity between clusters of texts.

Chapter 7, 'Register and modification in the noun phrase' (pp. 179–208), by Yolande Botha and Maryka van Zyl, examines the extent to which pre- and post-modifiers in the noun phrase distinguish among registers. To accomplish their objective, they do pairwise comparisons of ten registers – spoken, fiction, news, magazines, academic, interactive discussion, personal blog, news report/blog, frequently asked questions and historical article – using the pos-tagged versions of the *Corpus of Contemporary American English* (COCA) for the first five registers and the *Corpus of Online Registers of English* (CORE) for the last five. Given their research object, they normalize their data based on the number of nouns. To measure the differences between any two registers, they use the effect size metric and log likelihood significance tests. Results reveal that differences in frequencies of pre-modifiers mark the distinction between informational and oral-involved registers, whereas postnominal phrases and non-finite clauses distinguish between informational and narrative registers. Such results are in line with previous findings from multidimensional analysis studies; however, they have also come across significant differences between the register news, from COCA, and news report/blog, from CORE. The authors speculate that these differences are due to differing situational factors, as the news from CORE includes blog posts, but no definite answer can be given without further research.

The study reported in chapter 8, 'A register approach toward pop lyrics in EFL education' (pp. 209–34), by Valentin Werner, has a teaching application perspective and uses the multidimensional approach to reassess the assumed conversational profile of pop lyrics. The chapter carries out an additive multidimensional analysis to compare the frequencies of the linguistic variables in pop lyrics to those of the registers in Biber's (1988) seminal work, using the *Multidimensional Analysis Tagger* (MAT) to process a corpus of 1,842 songs by ninety-one different artists/bands. Valentin Werner's study finds apparently similar dimension means (in most dimensions) and derives analogous conclusions to Delfino's (2016) – who did an additive multidimensional analysis study of a corpus of 585 songs by The Beatles, Bon Jovi, Bruno Mars and Maroon 5, using the *Biber Tagger* and the *Biber TagCount*. The lyrics of songs in both works display characteristics of speech and writing due to the production circumstances pointed out by Werner: scripted, planned, monologic texts that are constrained by rhyme and are temporally, spatially and socially distant from their audience. Werner also posits that such status makes lyrics more than adequate to develop language awareness in learners, especially concerning the continuum characteristic of registers as well as the impact that contextual factors have on the 'actual linguistic layout of texts' (p. 230).

Like Valentin Werner's, the study by Tove Larsson, Magali Paquot and Douglas Biber in chapter 9, 'On the importance of register in learner writing: A multi-dimensional approach'

(pp. 235–58), was developed with a practical application in mind. It investigates the impact of register in learner writing against that of internal factors like first-language background. To this end, their study uses six corpora: *The International Corpus of Learner English* (ICLE), *The Varieties of English for Specific Purposes dAtabase* (VESPA), *The Louvain Corpus of Native English Essays* (LOCNESS), *The British Academic Written English Corpus* (BAWE), *The Michigan Corpus of Upper-Level Student Papers* (MICUSP) and *The Louvain Corpus of Research Articles* (LOCRA). The ICLE and the VESPA corpora represent learner writing, LOCNESS, BAWE and MICUSP provide samples of native-speaker student writing, and LOCRA samples of scientific articles written by experts. Such a variety of registers allows for comparisons across argumentative essays, research papers, scientific articles, native-speaker vs non-native-speaker writing and first-language background, as the sample of texts taken from ICLE and VESPA includes L1 speakers of French, Spanish, Norwegian, Swedish and Dutch. To carry out such comparisons, the authors apply the multidimensional analysis approach and identify two dimensions of variation, namely Personal vs Topic-focused Style and Evaluative Style vs Factual Descriptions. Results indicate that texts are mostly distinguished by their register, particularly in Dimension 1. Dimension 2 highlights the differences between expert and non-expert writers and shows that there is considerable variation in learners from distinct language backgrounds. To Larsson, Paquot and Biber the study has two major practical implications. First, the developers of English for Academic Purposes materials should be aware that learners from different language backgrounds have different language learning needs. Second, adequate reference corpora should be used in studies that look into the characteristics of learner argumentative texts, as texts written by experts are too different from them to be used as such.

Chapter 10, ‘Nominalizations in Early Modern English: A cross-register perspective’ (pp. 259–90), by Paula Rodríguez-Puente, looks at word-based nominalizations in Early Modern English. To do so, she traces nine Romance and native suffixes (*-age*, *-dom*, *-head*, *-hood*, *-ion*, *-ity*, *-ment*, *-ness* and *-ship*) in eighteen registers distributed along the formal–informal, speech–written continua. The texts used in her study were taken from three corpora, namely *A Corpus of English Dialogues* (CED), the *Penn–Helsinki Parsed Corpus of Early Modern English* (PPCEME) and the Early Modern English section of the *Corpus of Historical English Law Reports* (CHELAR). The nouns formed with one of the nine suffixes were identified with the help of *WordSmith Tools*. Rodríguez-Puente finds that the frequency of the suffixes used in the study is highly dependent on registers. Informal speech-related registers, such as diaries, drama, private letters, trial proceedings and witness depositions, have the lowest frequencies of nominalizations, whereas formal written registers like legal documents and public letters have the highest. Her results also show that narrative texts present much lower frequencies of nominalizations than persuasive documents. In terms of diachronic distribution, the frequency of nominalizations increases in most registers from 1560 to 1760, a period that also witnesses an increase in token and type frequency and ‘type richness’ (p. 283). The suffixes that showed the sharpest increase are the Romance suffixes, especially *-ion*, and the native suffix *-ness*. Such tendencies are not present in

two registers: trial proceedings and private letters. In the case of trial proceedings, token frequencies decrease and the suffixes lose productivity overtime. For private letters the scenario is slightly different, as token frequencies remain stable, and the suffixes do not lose so much productivity across the years.

In chapter 11, ‘Measuring informativity: The rise of compounds as informationally dense structures in 20th-century Scientific English’ (pp. 291–312), Stephania Degaetano-Ortlieb aims to measure the informativity of alternative phrasal structures – namely informationally dense compounds (sequences of two or more nouns) vs longer, less dense variants, such as prepositional phrases – that accounted for change in language use in scientific English in the twentieth century. To achieve such an aim, the author framed her study within Information Theory. The corpus selected is a section of the *Royal Society Corpus* (RSC), more specifically the section called *Proceedings A*, composed of texts written between 1905 and 1996 and belonging to mathematical, physical and engineering sciences. To verify whether the texts in the corpus became more informationally dense overtime, she applied data-driven periodization with Kullback-Leibler divergence, as such a technique allows for the identification of the decades in which the changes occur. Degaetano-Ortlieb offers three important findings: (1) an increase in the use of compounds overtime; (2) the relationship between higher informativity within the noun phrase and specialization, and lower informativity in the noun phrase and standardization; and (3) the retention of items with higher informativity and the replacement of items with lower informativity by compact language, such as compounds, across time. The author also situates a pronounced increase in two-noun compounds in the 1920s and shows a steady rise of three noun compounds, which peaked from the 1950s to the 1980s.

Last but not least, in the final chapter, ‘Exploring sub-register variation in Victorian newspapers: Evidence from the *British Library Newspaper Database*’ (pp. 313–38), Turo Hiltunen examines sub-register variation in nineteenth-century newspapers as well as the value of the *British Library Newspaper Database* for register studies, discussing both the operationalization of the notion of (sub-)register and the inter-dependency between sampling criteria, findings and interpretations. To this end, the author selects eight variables from Biber’s (1988) relevant dimensions – more specifically Dimensions 1 (Involved vs Informational Production), 2 (Narrative vs Non-narrative Concerns), 4 (Overt Expression of Persuasion) and 5 (Abstract vs Non-abstract Information) – namely private and suasive verbs, first-, second- and third-person pronouns, past tense forms, infinitives, conjuncts and one extra variable that will help him identify people’s names: sequences of two proper nouns. He then computes their frequencies for each text in the corpus. To make up for the lack of balance in the dataset, Hiltunen designs and tests different approaches to both compute frequencies and identify systematic patterns of variation across sub-registers. Results suggest that ‘there are systematic and linguistically meaningful patterns of variation across the sub-registers’ investigated (p. 333). Therefore, if researchers address the extant lack of balance between the sub-registers, the low quality of some texts and the



lack of information on some of the sub-registers, the *British Library Newspaper Database* is a good source of data for register studies.

In summary, the collection of chapters in this volume showcases the abundance of theoretical frameworks and methods that can be used to investigate, explore and describe different registers as well as challenge what we already know about such a construct. The top-down and bottom-up approaches presented here certainly inspire further research and pave the way for the discovery of new knowledge on register and register variation.

*Reviewer's address:*

Departamento de Letras  
Universidade Federal de São Paulo  
Rua Capital Federal, 550 ap. 51  
São Paulo, SP 01259-010  
Brazil  
[marcia.veirano@unifesp.br](mailto:marcia.veirano@unifesp.br)


#### References

- Berry, Margaret. 1995. Thematic options and success in writing. In Mohsen Ghadessy (ed.), *Thematic development in English texts*, 55–84. London: Pinter.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Delfino, Maria Claudia Nunes. 2016. O uso de música para o ensino de inglês como língua estrangeira em um ambiente baseado em corpus [Pop songs in the EFL classroom: A corpus-based approach]. Master's dissertation, São Paulo Catholic University.
- Halliday, Michael A. K. & Christian M. I. M. Matthiessen. 2014. *Halliday's introduction to Functional Grammar*, 4th edn. London: Routledge.
- Matthiessen, Christian M. I. M. 2019. Register in systemic functional linguistics. *Register Studies* 1 (1), 10–41.

(Received 30 November 2022)

doi:[10.1017/S1360674322000491](https://doi.org/10.1017/S1360674322000491)

**Isabel Moskowich, Inés Lareo and Gonzalo Camiña (eds.)**, *'All families and genera': Exploring the Corpus of English Life Sciences Texts*. Amsterdam and Philadelphia: John Benjamins, 2021. Pp. xv + 310. ISBN 9789027209245 (Hb), 9789027259622 (e-book).

Reviewed by Elisabetta Lonati , Università degli Studi del Piemonte Orientale, 'Amedeo Avogadro'

The volume *'All families and genera': Exploring the Corpus of English Life Sciences Texts*, edited by Isabel Moskowich, Inés Lareo and Gonzalo Camiña, is a very