

## Review

doi:[10.1017/S1360674324000133](https://doi.org/10.1017/S1360674324000133)

**Benedikt Szmrecsanyi and Jason Grafmiller**, *Comparative variation analysis: Grammatical alternations in World Englishes* (Studies in Language Variation and Change). Cambridge: Cambridge University Press, 2023. Pp. xvii + 218. ISBN 9781108491563.

Reviewed by Iván Tamaredo , Complutense University of Madrid

*Comparative Variation Analysis*, by Benedikt Szmrecsanyi and Jason Grafmiller, condenses the findings of a line of research initiated by Szmrecsanyi *et al.* (2016), which combines the main assumption of the probabilistic grammar framework that grammatical knowledge has a probabilistic component shaped by speakers' linguistic experience (e.g. Bresnan 2007) with work on postcolonial varieties of English (e.g. Schneider 2007). The book has two overarching goals, one theoretical and one methodological. The theoretical objective of the monograph consists in 'understand[ing] the plasticity of probabilistic knowledge of English grammar, on the part of language users with diverse regional and cultural backgrounds' (p. 1). In other words, the book's main theoretical contribution is to advance our knowledge of the extent to which probabilistic grammars vary across different dialects of English. From a methodological perspective, the monograph puts forward quantitative methods to analyze language variation in a rigorous and sound manner. To fulfil these two goals, the authors focus on three well-known grammatical alternations – the genitive, dative and particle placement alternation – in nine native and non-native varieties of English with different cultural and sociolinguistic backgrounds (list provided below). The data for the studies discussed in the book are gathered using different data sources and collection procedures, including both corpus-based and experimental approaches. The resulting datasets are analyzed using a series of state-of-the-art predictive modeling and dimensionality reduction statistical techniques, among others, which allow the authors to conclude that probabilistic grammars are, on the whole, fairly stable across varieties of English and that, whenever we find differences in the probabilistic grammars of varieties, these tend to derive mainly from differences in the effect strengths of (some) probabilistic constraints on grammatical variation. Moreover, there is on many occasions a division between Inner Circle and Outer Circle varieties, with the former being more homogeneous than the latter. Methodologically, the results of corpus-based and experimental approaches do not always converge, a finding that has important implications for studies in variationist linguistics. *Comparative Variation Analysis*, therefore, contains cutting-edge research at the crossroads of, to

---

name a few fields, variationist linguistics, probabilistic grammar, comparative sociolinguistics and dialect typology.

The monograph is divided into eight chapters. Chapters 1 and 8 introduce and conclude the book, respectively. Chapter 1 sets the scene by discussing the aims and research questions of the monograph, together with the data and methods. In addition, the authors provide a summary of the main findings and an overview of the theoretical framework of the book, including issues related to the fields of variationist linguistics, comparative linguistics, dialectology, dialectometry, dialect typology, probabilistic linguistics, psycholinguistics and English as a world language. Chapters 2 and 3 provide the theoretical framework of the studies discussed in later chapters. The data, the corpora used for the corpus-based studies and the probabilistic constraints on grammatical variation are described and explained in chapter 4. Chapters 5 and 6 present the results of the corpus-based analyses, with chapter 7 focusing on the experimental study.

Chapter 2 is the first of the two chapters devoted to the review of the relevant specialized literature, titled ‘Grammatical and syntactic variation’ (pp. 12–33). It focuses on grammatical alternations, which are approached from the perspective of traditional dialectology and modern variationist linguistics. Grammatical alternations, a term that refers to ‘alternate ways of saying “the same” thing’ (Labov 1972: 188) in the realm of morphosyntax, have been a controversial topic, as scholars have even debated whether such alternations exist. Section 2.3 provides a list of well-known grammatical alternations in English, the language under study in the monograph and the focus of this section. The section ends with a classification of grammatical alternations into three types, namely permutation alternations, in which the variants exhibit different word orders; insertion/deletion alternations, where a (grammatical) marker is either omitted or retained; and substitution alternations, in which a functional word/pattern may be replaced by another (see De Troij *in preparation*). The chapter moves to a review of previous comparative work on grammatical variation in English(es) from the perspective of probabilistic grammar (section 2.4) to then focus on the three alternations that are investigated in the monograph (section 2.5): the genitive, dative and particle placement alternation. Previous studies investigating each of the grammatical alternations are reviewed and a series of probabilistic constraints influencing variation are identified and briefly explained. The authors clarify their choice of alternations on the basis of practical issues: these three alternations have been extensively studied, the factors influencing variant choice are well understood, and they share several probabilistic constraints such as animacy or persistence effects. However, as the authors themselves acknowledge, they are all essentially permutation alternations, which limits the scope of the study: it would have been interesting to see as well the examination of other phenomena that constituted clear cases of substitution and insertion/deletion alternations, to assess whether different types of alternations behave differently.

Chapter 3 constitutes the second literature review of the book, in this case focused on the fields of World Englishes and dialect typology (pp. 34–55). The chapter begins with

an overview of the main models of World Englishes, in particular Kachru's (1985) Three Circles Model and Schneider's (2007) Dynamic Model. The perspective from dialect typology is also given, which focuses on the identification of structural features common to many dialects of the same language. Two of these features play an important role in the monograph, namely ANGLOVERSALS, features shared by all or most dialects of English, and VARIOVERSALS, features shared by a set of varieties with similar historical and sociolinguistic backgrounds. In the dialect typology literature, varieties are classified according to their variety type into native L1 varieties, indigenized L2 varieties and English-based pidgins and creoles, and according to their degree of language contact into high or low contact varieties. The nine varieties examined, namely British English (BrE), Canadian English (CanE), Irish English (IrE), New Zealand English (NZE), Hong Kong English (HKE), Indian English (IndE), Jamaican English (JamE), Philippines English (PhilE) and Singapore English (SgE), are then classified according to these models, which results in a representative sample of circles, phases and types.

Methodological considerations are discussed in chapter 4 ('The data', pp. 56–81). The chapter begins by describing the different types of data in the field of variationist linguistics. A distinction is made between observational and experimental data and, within the former, between the Labovian approach to data collection (i.e. through interviews and observation) and corpus-based approaches. Chapter 4 continues by providing, in section 4.2, an overview of the corpora the authors used to gather the data for the corpus-based studies. Essentially, the authors resorted to the two most-used corpora in the field of World Englishes, the *International Corpus of English* (ICE) and the *Corpus of Global Web-based English* (GloWbE). Following standard variationist practice, the variable contexts of the three grammatical alternations under study are defined in section 4.3, that is, the criteria employed to identify cases of the alternations in which the variants are truly interchangeable. The chapter ends with an explanation of the probabilistic constraints that the data were annotated for, which include the following, shared by all three alternations: animacy, definiteness, NP type, givenness, constituent length, priming/persistence and frequency (of lexical items in different slots of the constructions). In addition to these shared factors, the alternations were annotated for determinants of variation specific to each of them. In the case of genitives, the presence/absence of a final sibilant was also coded for, as well as the concrete semantic relation between possessor and possessum. The particle placement alternation was also annotated for the presence/absence of a postmodifying directional PP, the degree of semantic compositionality of each verb–particle combination, and surprisal, which refers to how predictable a verb is considering a particle and vice versa. No additional variables were annotated for in the case of the dative alternation.

The first part of the results is presented and discussed in chapter 5, titled 'Alternation-by-alternation analysis' (pp. 82–111). In the words of the authors, this chapter provides a 'jeweler's eye perspective' (p. 112) by analyzing each alternation individually and providing a detailed examination of the effects of the probabilistic constraints and their interactions on variant choice. To this purpose, they employ

state-of-the-art descriptive and inferential statistics, including variant rates or percentages, mixed-effects binary logistic regressions and random forests. These analyses allow the authors to conclude that the Inner/Outer Circle distinction is relevant regarding variant rates in all three alternations, with the particle placement alternation being especially susceptible to this distinction: the *s*-genitive variant, the ditransitive dative variant and the split particle variant are all significantly more frequent in Inner Circle varieties. In the three alternations, the multifactorial statistical tests uncover a high degree of homogeneity between varieties regarding which constraints are significant and the directions of their effects. This means that if constraint X is significant in variety A, it is highly likely that it will also have a significant effect in variety B and that the direction of the effect will be the same in both varieties; thus, the same variant will be (dis)favored in the same contexts. Differences exist, however, regarding the size or strength of these effects: even if one constraint is significant in all varieties and the direction of the effect is also the same, there are differences as to how strongly each variant is (dis)favored in each context. Based on these probabilistic fluctuations, the authors claim that different probabilistic grammars can be recognized in the data. In particular, the results of the analyses suggest that, in the case of the genitive alternation, different probabilistic grammars can be identified for HKE, PhLE, SgE, CanE and BrE. In addition, IndE and BrE seem to also exhibit different grammars for the dative alternation, and a distinction can be made between Inner and Outer Circle probabilistic grammars in the case of the particle placement alternation. As concerns differences between alternations, particle placement exhibits more regional probabilistic differences than the other two alternations or, in other words, it is more vulnerable to ‘probabilistic indigenization’ (Szmrecsanyi *et al.* 2016: 133), a finding that the authors explain based on the alternations’ degree of lexical specificity.

If chapter 5 adopted a jeweler’s eye perspective on the alternations, chapter 6 provides a bird’s eye view in ‘Distances, similarities, and coherence’ (pp. 112–40). The authors propose a new approach to examine differences in probabilistic grammar across dialects called Variation-Based Distance and Similarity Modeling (VADIS). This method, inspired by comparative sociolinguistics and quantitative dialectometry, examines varieties along three lines of evidence, namely statistical significance, effect strength and constraint ranking, to detect (dis)similarities in their probabilistic grammars. To this purpose, VADIS employs the output of per-variety mixed-effects regression and random forest models computed for each of the alternations to calculate a series of similarity coefficients that reflect how (dis)similar the probabilistic grammars of the varieties are. The main finding is that, overall, probabilistic grammars are stable across varieties. This stability, however, differs depending on the subset of data analyzed. Like the findings of chapter 5, the VADIS method identifies the particle placement alternation as the one more prone to probabilistic indigenization. In addition, by comparing the degree of stability across dialects of probabilistic grammars in spoken versus written language and in Inner versus Outer Circle varieties, the results show that spoken language and Outer Circle varieties are more heterogeneous. The authors also interpret the (dis)similarities obtained through VADIS as measures of

coherence. The genitive and particle placement alternations overlap substantially, which means that those varieties with similar probabilistic grammars regarding the genitive alternation also exhibit similar grammars for the particle placement alternation; the dative alternation is the odd one out. A similar conclusion can be extracted from the degree of coherence between spoken and written language: varieties with similar probabilistic grammars in speech also have similar grammars in written language. Finally, there is also a certain degree of overlap between lines of evidence, which implies that the three lines in VADIS measure different aspects of variation while simultaneously providing compatible results. Given the novelty of VADIS, chapter 6 ends by conducting a simulation to test the validity and reliability of the method. The authors apply VADIS to a constructed dataset containing five hypothetical varieties exhibiting different degrees of grammatical (dis)similarity. The results of the simulation suggest that VADIS is in fact able to accurately detect differences in probabilistic grammar.

The empirical component of the book is closed by chapter 7, which presents and discusses the results of the experimental study ('Experimental corroboration', pp. 141–65). On this occasion, the focus is only on the particle placement alternation in four varieties (BrE, NZE, IndE and SgE), and exclusively on one probabilistic constraint, namely the length of the direct object. An offline judgment task was used to estimate the participants' preference for each variant in different contexts of use reflecting different direct object lengths. In particular, the participants had to assign a percentage (out of a hundred) to each variant in each context depending on their preferences. The results of the experiment showed a strong correlation with the predictions of the corpus-based model, which means that the participants' preferences matched to a large extent the model's predictions of the probability of each variant in each context of use. However, differences between participants' preferences in the experiment and the predictions of the corpus-based model were also evident, for which the authors provide several potential explanations.

Chapter 8 concludes the monograph with an overall discussion of the results and their main theoretical and methodological implications, in 'Where are we now, and where to next?' (pp. 166–90). It starts by providing a detailed summary of the results of all the studies in the book and then proceeds to interpret them on the basis of the theoretical background. Szmrecsanyi and Grafmiller argue that the stable effect directions of the probabilistic constraints examined can be considered probabilistic angloversals, that is, features shared by all or most varieties of English at a probabilistic level. Similarly, similar effect strengths are suggestive of the existence of probabilistic varioversals, probabilistic features shared by varieties of English with similar historical, cultural and sociolinguistic backgrounds. Two main explanations are proposed by the authors for the existence of probabilistic universals: 'shared histories and shared humanity' (p. 173). Varieties of English exhibit similar probabilistic grammars, first, because they all have a common ancestry in that they have all developed from the same language and underlying grammatical system. In addition, the shared humanity perspective argues that all dialects are subject to the same universal production and comprehension biases originating in the cognitive architecture shared by all speakers. Considering

these two factors, it is not entirely surprising that universals should also exist across varieties of English at the level of probabilistic grammar. However, the results have, on many occasions, also showed a split between Inner and Outer Circle varieties, with the latter being more heterogeneous, that is, exhibiting more probabilistic differences, than the former. Inner Circle varieties are, overall, fairly homogeneous because of, again, their shared history, but also their shared mode of transmission: Inner Circle varieties are all acquired as L1s. Outer Circle varieties, on the contrary, are, or were at some point in their history, acquired mainly as L2s, which makes these varieties vulnerable to influences from substrate languages. The authors then review studies on the genitive, dative and particle placement alternations in dialects of English to identify possible substrate effects that may explain (some of) the differences between Outer Circle varieties uncovered by the analyses. Finally, the authors argue that the existence of probabilistic universals does not entail that variation is restricted to a certain type of constraints. The results of the different studies conducted showed that all types of constraints are subject to variation, even those claimed to be grounded on psycholinguistic biases. However, psycholinguistically grounded constraints do constitute the best candidates for angloversals, thus being more stable across varieties. Varioversals, on the other hand, exist when the effects of other constraints align in a group of varieties.

*Comparative Variation Analysis* is an excellent contribution to the study of language variation and will undoubtedly become a basic reading in the field. It contains cutting-edge research at the crossroads of many different areas of variation studies, making it relevant for all of them. This is not a coincidence, since the authors themselves mention that ‘one of the aims of the book is to cross-pollinate different research tracks in variation studies’ (p. 3), a goal that Szmrecsanyi and Grafmiller have achieved masterfully. The monograph can be considered a breakthrough from different perspectives. First, methodologically it puts forward a new comparative corpus-based method to assess the degree of probabilistic (dis)similarity between varieties of the same language using state-of-the-art statistical techniques and approaches. Second, it analyzes and compares both corpus-based and experimental data on the same linguistic phenomena. The results obtained are extremely detailed and complex, but the reader is more than able to understand and follow them thanks to the clear explanations of the authors and, to a great extent, their direct and straightforward writing style. Given the groundbreaking character of the book, some of the theoretical discussion of the findings is speculative. However, this is by no means a shortcoming of the monograph: as all good research does, *Comparative Variation Analysis* raises more questions than it answers, thus pointing to the exciting ‘road ahead’ (p. 190) in variation studies.

*Reviewer's address:*

*Department of English Studies: Linguistics and Literature*

*Faculty of Philology*

*Complutense University of Madrid*

*Plaza Menéndez Pelayo s/n*

---

28040 Madrid

Spain

[ivan.tamaredo@ucm.es](mailto:ivan.tamaredo@ucm.es)

### References

- Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base*, 75–96. Berlin: Mouton de Gruyter.
- De Troij, Robert. In preparation. North and south, bottom to top. Using big data to model syntactic variation in Belgian and Netherlandic Dutch. PhD dissertation, KU Leuven.
- Kachru, Braj B. 1985. Standards, codification and sociolinguistic realism: The English language in the Outer Circle. In Randolph Quirk & Henry G. Widdowson (eds.), *English in the world: Teaching and learning the language and literatures*, 11–30. Cambridge: Cambridge University Press.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Schneider, Edgar. 2007. *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.
- Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37(2), 109–37.

(Received 10 April 2024)