

## Book Review

## Keywords:

grammatical alteration; world Englishes; probabilistic grammar; variationist linguistics; multivariate analysis

## Corresponding author:

Bin Shao;  
email: [binshao@zju.edu.cn](mailto:binshao@zju.edu.cn)

## Comparative variation analysis at cross roads: perspectives and methods

Benedikt Szmrecsanyi and Jason Grafmiller, *Comparative Variation Analysis: Grammatical Alternations in World Englishes*

Cambridge; New York, NY: Cambridge University Press, 2023. Pp. xvii +218. Hardback \$125, ISBN: 9781108491563

Jianhong Wu  and Bin Shao 

School of International Studies, Zhejiang University, Hangzhou, China

Variation research on World Englishes has witnessed significant advancements as research in this area represents a convergence of variationist linguistics and related subfields, including contrastive linguistics, dialectology, probabilistic linguistics, psycholinguistics and English as a world language. Interests from various fields have put the research at risk of being fragmented into different research communities with inadequate exchange. Against this backdrop, this monograph aims to integrate different research strands in English variation studies with a coherent focus. Drawing on Probabilistic Grammar approach, which posits that grammatical knowledge is experience-based and that determinants of (morpho)syntactic variation are multifactorial and probabilistic in nature, it investigates three grammatical alternations in nine varieties of English within the English worldwide paradigm (Schneider 2007). The three alternations, namely the genitive, dative and particle placement alternations, are chosen for their shared constraints and regional variation, as evidenced in previous studies.

In the introductory chapter, the authors provide a comprehensive description of key concepts and research orientations that justify the study's research questions and methods. The study emphasizes language-internal constraints from a corpus-based variationist linguistics (CVL) perspective, focusing on the consistency of these constraints, which can be observed through differences and similarities across language varieties. Furthermore, it incorporates both usage-based and rule-based models of grammar to explain variation, with probabilistic grammar framing this as the association of conventional rules with probabilities learned through experience. Additionally, it draws on methodologies from psycholinguistics to explore metalinguistic judgments.

The variability under consideration in this study is what van Hout and Muysken (2016) referred to as Type 3 variability, which refers to alternate ways of saying the 'same thing'. Chapter 2 begins with a discussion on whether this kind of variability exists in grammar, followed by the empirical evidence from variationist linguistic literature which suggests that grammatical alternations are plentiful in English. Following this, the genitive, dative and particle placement alternations, which allow language users to manipulate constituent ordering to mark focus, are introduced. The genitive alternation involves different ways of expressing possession, e.g. *the president's brother* and *the brother of the president*. The dative alternation refers to different ways of expressing the recipient of an action, e.g. *give the police a statement* and *give a statement to the police*. The particle placement alternation represents the variation in the position of particles, e.g. *look the words up* and *look up the words*. The three alternations are investigated because the constraints governing them reflect subtle nature of grammatical knowledge, and the highly shared constraints also facilitate comparative analysis across alternations.

Chapter 3 presents two of the most influential models of world Englishes: Kachru's (1985) Three Circles model and Schneider's (2007) Dynamic Model. The nine English varieties examined in the study – British, Canadian, Hong Kong, Indian, Irish, Jamaican, New Zealand, Philippines and Singapore English – are introduced within

the context of these frameworks. Despite their limitations, both models offer valuable perspectives for interpreting research findings. The cross-varietal generalizations that are most relevant to the present study are features that tend to recur in varieties of a specific language and features shared by language varieties with similar socio-history, history depth, and modes of acquisition. Based on these observations, the authors discuss the state of the field from the perspective of dialect typology. Three extra-linguistic dimensions associated with English dialects are identified: variety type, exposure to contact, and geographical proximity, indicating the role of sociolinguistic setting in probabilistic indigenization.

In Chapter 4, the authors discuss the data types in different lines of variationist tradition, including small-sized naturalistic production data in the Language Variation and Change (LVC) paradigm, larger and more generalized data focusing on register, variety and language change in the CVL paradigm, and experimental data elicited from different tasks, demonstrating what corpus linguistic and experimental data can bring to traditional variationist research. The data for the study was collected from the International Corpus of English (ICE) and the Corpus of Global Web-based English (GloWbE). The authors maintain that the two corpora provide appropriate data source for comparative analysis at this scale. Procedures for collecting, filtering and annotating the datasets are also presented in detail in this chapter.

Chapters 5 and 6 analyze the corpus data through two complementary approaches. The authors first make generalizations with multivariate analysis techniques, and the effect directions manifest that probabilistic grammars are stable across varieties. However, quantitative differences exist in the strength of effect on variation. The three alternations differ in how much they are affected by regional difference, which can be attributed to the numerous social, cognitive and linguistic factors that shape the development of regionally specific probabilistic grammars. In Chapter 6, Variation-Based Distance and Similarity Modeling (VADIS) is employed to quantify the distance and coherence among English varieties with regard to how language users choose different ways of saying the same thing in the three alternations. In terms of similarity, the Inner Circle varieties form a tighter typological cluster than the Outer circle varieties. Distance-Based Coherence (DBC) is measured using sub-datasets from three aspects, that is, the linguistic distance between varieties, between spoken and written genre, and between lines of evidences in comparative sociolinguistics. These measurements provide a broader perspective that complements the close-up analysis presented earlier.

Chapter 7 presents experiments designed to corroborate the scope of patterns observed in corpus models. A series of rating tasks are administered to check whether certain probabilistic constraints are consistent in corpus-based research and users' introspective preferences in particle placement alternation. The outcome shows strong evidence for the capacity of corpus data to model probabilistic knowledge, and participants' sensitivity to contextual features, specifically the length of the direct object in this case.

However, corpus data and experiments yield diverging conclusions regarding the preference for length effects across varieties. The authors attribute this to the difference in the nature of production and comprehension tasks, the ease of detecting internal processing-based factors in comprehension tasks compared with production tasks, and language change that may occur in the time span of corpus data collection.

Chapter 8 concludes this monograph by reviewing the research findings and discussing the implications of the study. The authors propose that common historical patterns and structural effects of L1 and L2 acquisition are major factors contributing to typological divisions in language varieties, and they also call for attention to the role of internal linguistic factors, social-stylistic factors and cognitive factors, and how these factors interact in language variation and change.

The research project exemplifies how comparative variationist studies can benefit from the ideas and perspectives in related subfields of linguistics. The combined use of large-scale corpus data and experimental analysis serves the ultimate goal of examining the stability and plasticity of probabilistic grammar across English varieties around the globe. The VADIS method has been proven to be effective in integrating the separate branches of variation studies. One advantage of VADIS lies in its ability to compare different variation phenomena in dozens of varieties at a time. Research findings have become much more accessible with supplementary visualization techniques such as Multidimensional Scaling (MDS). Furthermore, the prospect of its application in the analysis of diverse varieties and levels of language structures is exciting. For instance, it can be employed to measure probabilistic stability across time and registers, or to examine lexical and phonological variables. We also see the enthusiasm for embracing methodological pluralism, as manifested by the experimental corroboration in Chapter 7. The probabilistic nature of grammar is inherently related to how language is produced and processed. Experimental methods allow researchers to examine covariation and probabilistic conditioning in production and comprehension, and they are relatively easy to replicate compared to gathering large-scale corpus data. What's more, the possible diverging evidence from corpus and experimental results may reflect different aspects of the variation, thereby prompting the exploration of new avenues for future research. The comparative analysis in the study aims to examine the consistency of probabilistic grammars regulating variant choices across varieties, which can be regarded as evidence for the existence of a "common core" (Quirk et al. 1985) across world English varieties. Though probabilistic grammars of the alternations under study are stable across varieties, it is worth a second thought before making typological generalizations that the constraints governing the alternations are part of the 'common core' of standard English, as the VADIS coefficients are based on a combination of measures.

One minor flaw that would raise doubt in the study may be the validity and reliability of VADIS. On the one hand, researchers should carefully choose high-quality and

representative corpus data for analysis. On the other hand, external validation of VADIS can be achieved by comparing its results with experimental data and by checking its correlation with other independent data about the language patterns under study. It is not possible to explain probabilistic indigenization and variation without taking into account diachronic factors. Based on synchronic datasets in the study, future studies can focus on historical developments of the varieties. Additionally, the acquisition of the alternations among L1 and L2 users may also provide insight into the explanation of variation. Overall, this monograph provides a comprehensive and coherent review of literature, offering a quick access to the field of comparative variationist linguistics for beginners, and the sophisticated analytical methods and broad scope of this volume make it a must-read for researchers.

### References

- Kachru, Braj B. 1985. "Standards, Codification and Sociolinguistic Realism: The English Language in the Outer Circle." In *English in the World: Teaching and Learning the Language and Literatures*, edited by Henry G. Widdowson, 11–30. Cambridge: Cambridge University Press.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Schneider, Edgar. 2007. *Postcolonial English: Varieties Around the World*. Cambridge: Cambridge University Press.
- Van Hout, Roeland, and Pieter Muysken. 2016. "Taming Chaos. Chance and Variability in the Language Sciences." In *The Challenge of Chance*, edited by Klaas Landsman and Ellen van Wolde, 249–266. Cham: Springer International Publishing.



**JIANHONG WU** is a PhD candidate at Zhejiang University. Her research interests include corpus linguistics, cognitive linguistics, pragmatics, and language variation and change. She has published a number of research articles in linguistics. Email: [jianhongwu@zju.edu.cn](mailto:jianhongwu@zju.edu.cn)



**BIN SHAO** is a professor at the School of International Studies, Zhejiang University. His research interests include corpus linguistics, English lexicology, cognitive linguistics, and language variation and change. He has published articles in journals including *Journal of English Linguistics*, *Applied Linguistics*, *Corpus Linguistics* and *Linguistic Theory* and *Lingua*. Email: [binshao@zju.edu.cn](mailto:binshao@zju.edu.cn)