Hunston, Susan. 2000. Evaluation and the planes of discourse: Status and value in persuasive texts. In Susan Hunston & Geoff Thompson (eds.), *Evaluation in text: Authorial stance and the construction of discourse*, 176–207. Oxford: Oxford University Press.

Hunston, Susan & Hang Su. 2017. Patterns, constructions, and local grammar: A case study of 'evaluation'. *Applied Linguistics* 40(4), 567–93.

Hyland, Ken & Feng (Kevin) Jiang. 2018. 'We believe that …': Changes in an academic stance marker. *Australian Journal of Linguistics* 38(2), 139–61.

Hyland, Ken & Polly Tse. 2005a. Evaluative *that* constructions: Signalling stance in research abstracts. *Functions of Language* 12(1), 39–63.

Hyland, Ken & Polly Tse. 2005b. Hooking the reader: A corpus study of evaluative *that* in abstracts. *English for Specific Purposes* 24(2), 123–39.

Kohnen, Thomas. 2009. Historical corpus pragmatics: Focus on speech acts and texts. In Andreas H. Jucker, Daniel Schreier & Marianne Hundt (eds.), *Corpora: Pragmatics and discourse*, 11–36. Amsterdam: Brill.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Stephen Pinker, Martin A. Nowak & Erez Lieberman Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014), 176–82.

Moskowich, Isabel, Inés Lareo & Gonzalo Camiña (eds.). 2021. *'All families and genera': Exploring the Corpus of English Life Sciences Texts*. Amsterdam: John Benjamins.

Romero-Trillo, Jesús 2017. Editorial. *Corpus Pragmatics* 1, 1–2.

Sinclair, John. 1986. Fictional worlds. In Malcolm Coulthard (ed.), *Talking about text: Studies presented to David Brazil on his retirement*, 43–60. Birmingham: English Language Research.

Taavitsainen, Irma. 2011. Medical case reports and scientific thought-styles. *Revista de Lenguas para Fines Específicos* 17, 75–98.

Taavitsainen, Irma, Andreas H. Jucker & Jukka Tuominen. 2014. *Diachronic corpus pragmatics*. Amsterdam: John Benjamins.

Paula Rautionaho, **Hanna Parviainen**, **Mark Kaunisto and Arja Nurmi (eds.),** *Social and regional variation in World Englishes: Local and global perspectives* (Routledge Studies in Sociolinguistics). New York and London: Routledge, 2023. Pp. xiii + 219. ISBN 9781032130361.

Reviewed by Axel Bohmann 🄾, Universität Freiburg

Social and Regional Variation in World Englishes is a Festschrift for Juhani Klemola on the occasion of his 65th birthday. As typical of this publication format, and as indicated by the title, the volume comprises a topically diverse range of contributions. In addition to the unifying focus on variation and World Englishes, however, all chapters follow a corpus-based, empirical methodology, giving coherence to the book despite its breadth of topics. Following a foreword by Kate Corrigan, the first chapter is

the editors' general introduction to the volume (pp. 1–7). They highlight both the need for an integrationist perspective on variation in World Englishes, combining regional, stylistic, social and intra-linguistic factors, and the importance of advanced statistical techniques. Conscious of the heterogeneity of contributions, the chapter frames the volume as reflecting the current breadth of the field.

The first research contribution is Terttu Nevalainen's study of variation in first- and second-person possessive pronouns (*mine/thine* versus *my/thy*) between 1460 and 1650 (pp. 8–31). Nevalainen draws on her own data from the *Corpus of Early English Correspondence* (CEEC) as well as results from other studies using the *Helsinki Corpus*, to untangle the effects of style, register and social variation on an ongoing change. Whereas social variation is operationalized in conventional variationist terms with the variables gender, social rank and regional background of a writer, the notions of register and style are not always kept clearly distinct. In the first instance, Nevalainen treats the level of familiarity between correspondents in CEEC, binarily coded as intimate versus distant, as register. This essentially amounts to formality as the key criterion, which is very much in line with the original, Labovian definition of style. The supplementary data from the *Helsinki Corpus*, covering the range of text types represented therein, correspond much better with established notions of register as situationally conditioned variation. The study's strength lies in its careful treatment of the variable under discussion, identifying subtle language-internal constraints, as well as bringing heterogeneous datasets into conversation with each other. The key finding is a two-stage model for the effects of social and register variation in ongoing change. According to this model, as a change is incoming, it is predominantly subject to social stratification, whereas at later stages social variation evens out and gives room to register differentiation. The chapter thus provides a highly interesting framework to be explored in future empirical research.

Markku Filppula's contribution is by far the shortest in the volume (pp. 32–42). At nine pages excluding references, it is a corpus study of *have to* versus *have got to* in English and Irish English, including synchronic and historical corpora for both varieties as well as supplemental material from India and the Hebrides. The results conclusively show that *have got to* is both historically and synchronically better established in English English compared to Irish English, where it occurs as a minority variant with rates below 10 percent compared to *have to*. The chapter speculates on the specific contact situation in nineteenth-century Ireland and the mode of acquisition of English leading to preference of the structurally simpler variant. While this hypothesis is not implausible, the chapter does not provide any data to corroborate or falsify it. To argue more convincingly about divergent or common roots in the use of *have (got) to*, tools such as comparative sociolinguistic methods could have been helpful and added the empirical nuance to turn this interesting chapter into something more substantial.

Chapter 4 is the first to deliver on the editors' promise of showcasing advanced statistical techniques (pp. 43–65). Paula Rautionaho and Mark Kaunisto tackle recent, real-time change in the use of *was/were* with pronominal subjects *we*, *you* and *they* in British English. Drawing on the *British National Corpus* (BNC) and its successor, the

BNC14, they are able to compile a set of comparable data representing two sampling points: 1994 and 2014, respectively. In addition to this real-time dimension, social (gender, class, age and region) as well as intra-linguistic information (polarity, inversion, subject pronoun) is retained. The most striking finding is a drastic shift away from (non-standard) *was* in the relatively short period of twenty years. Against this general backdrop, all further variables add nuances. In particular, and this is where the authors draw on advanced statistical analysis, a generalized linear mixed-model tree analysis shows not only a decline in frequency of the non-standard variant, but a corresponding leveling of constraints. The biggest question these findings raise is the extent to which they may be an artifact of corpus composition rather than an actual change. The fact that the speaker cohort who were 0–29 years old in 1994 use *was* at a rate of 5.3 percent, but speakers aged 30–49 in 2014 (i.e. roughly the same generation of speakers) show a rate of only 0.6 percent, certainly requires explanation. The authors touch on the issue of corpus comparability, but it is ultimately one that deserves attention beyond the scope of this chapter.

The volume's focus on World Englishes, so far at best hinted at in Filppula's contribution, is further developed in the following chapters, starting with Raquel Romasanta's (pp. 66–90). Her study of complementation patterns with *regret* in five African varieties of English, using British and American English as reference varieties, is based on the *Corpus of Global Web-based English* (GloWbE). Romasanta articulates three transparent research questions regarding the relationship between the varieties studied in terms of their complementation behavior, namely, whether there is notable surface variability among the varieties, whether this variability also manifests itself in the system of constraints governing the variable, and whether geographical proximity explains observed similarities and differences. The data are 2,089 instances of the variable, coded for twelve language-internal predictor variables. Since the relationship between observations and number of variables is problematic for conventional linear modeling, the study constructs random forest analyses of these predictors, separately for each variety. The rankings of variable importances according to each random forest are then descriptively compared to each other. Clustering of the varieties, based on ratios of *regret*-complementation patterns per individual predictor, is also performed. The results are taken as potential support for several hypotheses, namely, simplification as a key process in less developed varieties of English, regional proximity as an influential factor, and even potential substrate effects. The trouble is that, despite the general methodological sophistication of the chapter, no clear operationalization of these hypotheses or the research questions articulated at the outset is given. If the interest is in the differences among varieties, for instance, it would seem more plausible to include variety as a predictor in one general model, rather than constructing separate by-variety models and then descriptively comparing them. As such, the chapter stimulates a range of theoretical perspectives on African varieties of English, but has difficulty addressing its questions with conclusiveness.

The chapter by Heli Paulasto and Lea Meriläinen widens the scope not only geographically, including varieties of English from Europe, Africa, North America and

Asia, but also in terms of variety type (pp. 91–122). The study explores patterns of preposition omission in English as a first language (ENL), English as a second language (ESL) and English as a foreign language (EFL) settings, employing no less than ten corpora of both written and spoken English. These represent writing from Swedish, Finnish, Chinese and Japanese learners, a written reference corpus of North American students' essays, as well as conversational data from Britain (the British segment of the *International Corpus of English* [ICE], the *Survey of English Dialects*, the *Welsh English Corpus*), Singapore and Nigeria. Being a zero form, preposition omission is difficult to automatically analyze and hence remains relatively understudied. The authors opt for careful, manual coding of 30,000 words in each of their corpora, yielding a total of 394 cases of omitted prepositions. Univariate analyses in the form of chi-square tests are run to compare the distribution of cases across context variables, such as whether the preposition is under the scope of a head word (e.g. *listen to*), whether it is part of an adverbial (e.g. *He lives in London*), whether it is a complex (e.g. *up to*) or simple preposition, and whether it is stranded (e.g. *the mood she was in*). Another factor, and one that strikes me as slightly problematic, is whether the preposition is optional in Standard English. A case in point is their example (27), the relevant part of which is 'to possibly prevent things Ø growing out of hand' (p. 111). The problem with this example is that it does not have to be read as containing an underlying preposition. Alternatively, 'things growing out of hand' may be read as the noun phrase complement to *prevent*. When dealing, as the authors do, with contact varieties, further issues arise. Example (2), from *ICE-Nigeria*, reads 'I know say Ø Lagos dey reach five hundred'. Here, the language is not English at all, but Nigerian Pidgin, and *say* acts as a complementizer, whereas *dey* is a preverbal aspect marker. Apart from assuming an underlying English preposition in a sentence that is not English, even the most plausible translation of this Pidgin example into English does not require a preposition: 'I know that Lagos is reaching five hundred'. These issues do not discredit the authors' work so much as illustrate how complex the problem of preposition omission is. The underlying level of complexity is indeed one of the key results stressed in the conclusion. Still, in the face of difficult data and heterogeneous tendencies, the study convincingly argues for a systematic difference in the contexts in which ENL/ESL varieties, on the one hand, and EFL varieties, on the other, omit prepositions. In the former, a limited set of recurring contexts, apparently entrenched at the community level, accounts for the bulk of omissions, whereas learner varieties are characterized by more spontaneous, less systematic omissions.

Paulasto and Meriläinen's study, with its focus on emergent patterns of omission, is nicely complemented by Robert Fuchs' study of colonial lag regarding the negative scalar conjunction *and that too* in Indian English (pp. 123–48). Rather than relying on complex statistical tools, the study makes informed use of a range of corpora to reconstruct diachronic developments in Indian, British and other varieties of English. Establishing historical relations of this kind is notoriously difficult, since diachronic corpora are notably lacking for many post-colonial varieties of English. However, triangulating between the British and Indian segments of ICE, the

Kolhapur Corpus of written Indian English, *Google Books* and GloWbE, Fuchs arrives at a surprisingly plausible conclusion. The chapter clearly shows that negative-scalar *and that too* is used in present-day Indian English but not present-day British English and that it was in use in the British input variety in colonial India. To rule out colonial revival as an explanation, the study demonstrates consistent frequencies of use between the sampling time of the *Kolhapur Corpus* (1978) and *ICE-India* (1990s). Further, results from GloWbE show that *and that too* is widespread in South Asian varieties of English in general, with Pakistani, Sri Lankan and Bangladeshi English showing significantly higher frequencies than all remaining varieties. The implications are explored in relation to historical processes, which, if speculative at times, paint a plausible picture of cross-varietal influence that invites further attention in the future. Fuchs' conclusion that the notion of colonial lag – pertaining to individual features rather than varieties on the whole – is far from a myth is certainly well supported by this case study.

   Like Fuchs' chapter, Patricia Ronan's contribution focuses on a single form, whose frequency distribution she traces across a number of large corpora (pp. 149–65). The study explores the emergence and spread of the apologetic and blame-taking expressive marker *my bad* over the past thirty years, with a focus on American English but also an eye towards global dissemination. Its diachronic rise and spread from informal to more formal genres can clearly be seen in the results from the *Corpus of Historical American English* (COHA) and the *Corpus of Contemporary American English* (COCA). I am less convinced about the role Ronan assigns the media in perpetuating use of *my bad*. It is true that genres representing telecinematic discourse lead the development, but the relationship between these genres and the feature's use in vernacular speech is difficult to determine with the corpora at hand. Neither COHA nor COCA includes texts that represent anything like spontaneous, unmonitored speech. The genre label 'spoken' in COCA relies exclusively on transcripts from TV and radio talk programs. Other features that are well attested as part of general spoken American English – for instance, *gonna*, *ain't* or *you guys* – are also between four (*you guys*) and twenty (*gonna*) times more frequent in the telecinematic data in COCA than in the 'spoken' part. It would be implausible to interpret each of these as a media innovation, and the same caution should be applied to *my bad*. At the level of theory, any argument about the media's role in language change would do well to incorporate recent discussions of this topic, as for instance prominently reflected in Sayers' (2014) 'mediated innovation model' and responses to it.

   Moving from telecinematic to social media, Mikko Laitinen and Masoud Fatemi's chapter is largely a methodological contribution (pp. 166–90). The authors identify absence of social-demographic information on Twitter as an obstacle in computational sociolinguistics. They specifically focus on the role of social network information. Drawing on openly available information from Twitter, Laitinen and Fatemi calculate five measures of network strength. Three of these are standard implementations in

network analysis, but the authors also propose two new ones. On the basis of these measures, they categorize a set of thirty-five randomly sampled Twitter networks, each built around a single user node, into a strong and a weak group and use this difference to predict the normalized frequencies of two linguistic variables. Against the relatively complex and informationally rich procedure of network strength estimation, the conflation of the data into a binary distinction at a relatively arbitrary cut-off appears a bit anti-climactic. This is especially true since no statistically significant differences emerge between the 'strong' and 'weak' networks with regards to either of the linguistic features. It is not entirely clear what motivated the authors to discretize the data rather than using each calculated network strength coefficient as a predictor. The strength of the chapter is that it introduces researchers to a range of network measures, which may find more productive application in future research.

The final chapter, written by Sebastian Hoffmann, Sabine Arndt-Lappe and Peter Uhrig, does not so much round off the volume as provide new and stimulating outlooks (pp. 191–213). In terms of investigated phenomena, it is the only contribution concerned with phonology. More precisely, the authors combine the Principle of Rhythmic Alternation (PCA), according to which stressed and unstressed syllables alternate in speech, with the typological distinction between stress- and syllable-timed languages. Their hypothesis is that in native varieties of English the PCA predicts avoidance of iambic clash, i.e. a word ending in a stressed syllable followed by a word beginning with another stressed syllable. In contact varieties of English with a set of syllable-timed substrate languages, the argument is that iambic clash should play a smaller role, if any. The innovation of the study lies in adopting a methodology originally designed for testing constraints in Optimality Theory for comparisons of different varieties of English. On the basis of the GloWbE corpus, the study demonstrates that stress-timed Inner-Circle varieties of English consistently place more emphasis on the avoidance of iambic clash than syllable-timed Outer-Circle varieties. The effect remains intact when controlling for the genre composition of GloWbE (blog versus general websites), giving the results additional credibility. The authors remain cautious about the generalizability of their findings and identify several possible confounds and further steps required to corroborate the results. This care is well taken, yet behind all necessary hedging the general findings appear both robust and worthy of further investigation.

On the whole, *Social and Regional Variation in World Englishes* includes many stimulating and well-executed studies. It provides a good overview of current, corpus-based approaches to variation in English worldwide. The coverage of corpora in particular is as exhaustive as one can expect between the covers of a single book, with evidence from twenty-nine different corpora being included. The fact that World Englishes are covered as a topic but the authors are exclusively from Europe is an imbalance no doubt owed to the Festschrift nature of the book and partly to the state of the field in general. One aspect which the editors emphasize in the introductory chapter but which I find difficult to reconcile with the sum of contributions is the focus on advanced statistical techniques. By no means all chapters draw on said methods, and

some of the most insightful ones make do entirely without them. Where sophisticated quantitative approaches are used, as in Romasanta's or Laitinen and Fatemi's chapters, the final analyses relapse into simple bivariate or descriptive techniques. These chapters are still worthwhile reads, but not necessarily for reasons of statistical sophistication. The volume's merit rests less in pushing any methodological or theoretical boundaries than in providing a showcase of current, corpus-based research into variation in English.

Reviewer's address:
Englisches Seminar
Universität Freiburg
D-79085 Freiburg i. Br.
Germany
axel.bohmann@anglistik.uni-freiburg.de

Reference

Sayers, Dave. 2014. The mediated innovation model: A framework for researching media influence in language change. *Journal of Sociolinguistics* 18(2), 185–212.

Olga Timofeeva, *Sociolinguistic variation in Old English: Records of communities and people* (Advances in Historical Sociolinguistics 13). Amsterdam and Philadelphia: John Benjamins, 2022. Pp. xvii + 204. ISBN 9789027211347.

Reviewed by J. Camilo Conde-Silvestre, University of Murcia

'Bad data' is one of the methodological difficulties that confront the practitioner of historical (socio)linguistics. This is due to the fragmentary nature of the available material – mainly written, formal texts (which excludes spoken and colloquial language) by members of privileged, literate communities (which, for most historical periods, excludes females and lower-class speakers) and isolated from social and communicative contexts – so that many registers, styles and the socio-demographic background of informants are absent (Hernández-Campoy & Schilling 2012: 65–73). For obvious reasons, this is in a direct proportional relationship to the impending weight of time depth: the further back in time the object of analysis, the greater the bad-data problem. Thus, sociolinguistically oriented research on the history of English has privileged the early and late modern periods, with a relative engagement with Middle English texts and a