# Lexicosemantic diffusion in World Englishes: variable meaning–form relations in prospective verbs[1]

E D G A R   W.   S C H N E I D E R
*University of Regensburg and National University of Singapore*
(Received 18 January 2023; revised 22 March 2023)

This article suggests that in the investigation of World Englishes, which has tended to focus on syntactic, phonological and lexical preferences, the analysis of shifts in word meanings (and meaning–form relations in lexical items) needs to be incorporated. Exemplary small-scale studies show that in polysemic words certain varieties come to prefer specific meanings, and in word fields some varieties begin to prefer certain forms over others. Based on analyses of different ICE corpora, a set of prospective verbs, their meaning relationships and their varying correlations with syntactic construction choices in different varieties are investigated quantitatively (using HCFA and conditional inference trees) and qualitatively (showcasing interesting innovative, possibly emerging uses in some countries). Regionality is consistently shown to be a weakly conditioning significant factor. Thus, it is suggested that lexicosemantic variability and diffusion in the evolution of World Englishes deserve and need to be investigated systematically.

**Keywords:** polysemy, prospective verbs, semantic change, word fields, World Englishes

## 1   Introduction: World Englishes and lexical semantics

As is well known in English linguistics, over the last few centuries a large number of new, stable and distinctive varieties of English have emerged around the globe, especially in Asia and Africa. Originally they were products of colonial expansion, 'daughter varieties' of (mostly) British or (rarely) American English. Throughout the latter part of the twentieth century these varieties were pushed by globalization, and especially in the twenty-first century they have been vigorously expanding. Investigating the histories, sociolinguistic and political settings and structural properties of these 'World Englishes' has grown to be a vibrant, fashionable and versatile research paradigm with

---

many different facets. A rich universe of research outlets and activities has been established, including two major journals (*English World-Wide*, *World Englishes*), several book series (e.g. Varieties of English Around the World, Routledge Studies in World Englishes, …), many handbooks (e.g. Schneider *et al.* 2004; Kirkpatrick 2010; Klemola, Filppula & Sharma 2017; Schreier, Hundt & Schneider 2020; Nelson, Proshina & Davis 2020), textbooks (e.g. Kirkpatrick 2007; Mesthrie & Bhatt 2008; Schneider 2020a), collective volumes and monographs, conference series, web resources, and many individual descriptions of specific features in specific varieties.

Research topics in the study of World Englishes have encompassed their sociohistorical growth and models of their evolution, issues of language pedagogy and language policies in multilingual settings, investigations of linguistic attitudes, of contact effects and mixing, and, above all, analyses of structural properties of these young varieties which have emerged in the process of 'structural nativization' (Schneider 2007; Hoffmann 2021). Distinctive properties of World Englishes are products of different kinds of processes (Schneider 2007: 97–112), notably continuity (features passed on from the donor variety across generations), innovation (newly evolved properties), and contact effects triggered by indigenous languages (e.g. lexical borrowing and structural transfer, possibly determined by specific constraints and 'filters'; cf. Bao 2015). Investigations have looked into many phenomena on the levels of phonetics/phonology, lexis and grammar (and, rarely but increasingly, pragmatics and culture).

Studies of semantic change have been a minor component in this endeavor, and have mostly looked into individual instances of word meaning change. Examples of such documentation and exemplification can be found, for example, in Schneider (2020a: 209–10), Wasserman (2020) and Mehl (2018), where some such studies are surveyed in section 2.3. There is a small set of studies that attempt to tackle processes of semantic change in a more principled fashion by checking what happens to a few select verbs in new varieties of English. Mehl (2018) investigates possible onomasiological alternates of the verbs *make* and *give* in three varieties and finds that unlike British and Singaporean English, which prefer the Germanic-derived monosyllabic target verbs over possible Latin-derived, polysyllabic alternates such as *produce* or *provide*, Hong Kong English lacks such style sensitivity and consistently prefers the simple verbs. Similarly, Werner & Mukherjee (2012) analyze the changing polysemy of *give* and *take* in British English as opposed to Indian and Sri Lankan English, and Bruckmaier (2017) offers a comprehensive documentation of syntactic–semantic variability in uses of the verb *get* in three different varieties (British, Singaporean and Jamaican English).

These studies are interesting and important but limited in scope, largely to examples of isolated words or small groups of lexical items. Clearly, broad and systematic research on lexicosemantic diffusion, basic patterns in the transmission of word meanings, is lacking. The fundamental research question 'What happens to meanings of words in the evolution of New Englishes?' has practically not been addressed so far in a principled fashion. In other words, in the process of the

emergence of New Englishes,[2] are meanings of words, just like lexemes, sounds, or structural patterns, passed on and modified in similar ways? Are they diffused and retained largely without major changes, or modified and altered systematically somehow? This is the question which inspires the present article, which may be seen as an initial step into a broader, novel research branch.

## 2 Lexicosemantic diffusion in World Englishes

Word meanings vary and change in the course of time, and they do so in a variety of systematic ways.[3] In the present context of investigating semantic diffusion and change, two perspectives are particularly relevant and interesting (see also Grondelaers, Speelman & Geeraerts 2010): (a) changes of the meanings of polysemic words (a semasiological perspective), and (b) changes of paradigmatic relationships within a 'semantic field', of how semantic space is subdivided among words with similar or overlapping meanings (an onomasiological perspective). Both types are illustrated in the next two sections (building on selective results from earlier work; Schneider 2020b).

### 2.1 Exemplary case studies I: polysemy

Most English words, especially verbs, are polysemic, i.e. they have varying meanings, or shades of meaning, which are usually listed, defined and counted in dictionaries. Polysemy is inherently 'prototype-based' (Grondelaers, Speelman & Geeraerts 2010: 988).[4] Context disambiguates, i.e. co-occurring lexeme types, collocations or syntactic constructions select and activate individual meanings, providing the decisive clues as to which of the possible meanings of a word is meant. See the examples in (1) and (2).

(1) Polysemy of *poor*
    (a)  *poor people* 'lacking property'
    (b)  *this poor guy* 'lacking happiness'
    (c)  *a poor performance* 'lacking quality'

(2) Polysemy of *consider*
    (a)  *He's considering moving.* + Ving: 'action of one's own'
    (b)  *Let us consider this problem.* + NP: 'think about systematically'
    (c)  *They consider themselves as liberals/liberal.* + NP *as* NP/ AdjP: qualifying, 'believe to be in a category / have a property'

---

[2] In line with much writing in the field I use 'World Englishes' as a cover term for all varieties globally and 'New Englishes' as the label for newly evolved second-language varieties, mostly in Asia and Africa (see Schneider 2020a: 30–1).

[3] A classic taxonomy of types of change was introduced by Stern (1931), and a modern, cognitively inspired one was proposed by Blank (1999). For more recent theoretical surveys of lexical semantics, lexicosemantic variation and change see Geeraerts, Grondelaers & Bakema (1994), Geeraerts (2009: esp. ch. 1), Grondelaers, Speelman & Geeraerts (2010) and Riemer (2016: part V).

[4] The classic analysis of a prototypical word meaning structure (of the verb *lie*) is Coleman & Kay (1981).

The adjective *poor* basically denotes a lack of something, with a level of some target value falling below an expected standard; but the understanding of what precisely is missing (which is part of the word meaning in context) depends upon the type of co-occurring noun, as shown in (2a–c). The verb *consider* refers to some sort of a thinking process. How this process is carried out and what is prototypically being reflected on is signaled by the syntactic complementation structure: with a verbal *-ing* form following (and typically the verb itself also being in the progressive form) a subject thinks rather light-heartedly about the possibility of carrying out some future action oneself (2a); a noun phrase object identifies an abstract object of reflection and implies that this reflection is done rather seriously and systematically, weighing various aspects (2b); and a complex transitive verb complementation (with or without *as* between an object and object complement) means a static opinion is held as to which category the object belongs to or which property it has (Schneider 1988a, II: 59–70, 347–8).

Since meanings of words can only be indirectly deduced and context provides the decisive clues, changes of meanings in variety evolution need to build upon what can be observed on the surface, i.e. syntactic or lexical context. The following example illustrates this by documenting shifting meanings of the verb *learn*.

(3)  Three basic meanings of the verb *learn*
   (a)  'receive information', e.g. *that night I learned that I have a skill ...* (ICE-HK s2a-037.txt) [+ *that*/*about*/*wh**]
   (b)  'store information', e.g. *he learned a number of Malay words* (ICE-Sing s2a-066.txt) [+ NP]
   (c)  'acquire ability to do sth.', e.g. *I learned to cultivate an ego* (ICE-Ind, w2f-012.txt) [+ *to*-infinitive clause]

The examples in (3) illustrate what I posit to be the three basic meanings of this verb, together with their characteristic syntactic patterns and the sources of the examples (all from Asian Englishes): *learning* may denote simply receiving information inadvertently, e.g. hearing it accidentally, and this information is typically expressed by a finite object clause (connected with *that* or a *wh*-form) or a noun phrase connected with *about*, as in (2a); it may mean that some mental object, typically expressed as a noun phrase, is deliberately memorized, stored in mind, as in (2b); and it may relate to the acquisition of the ability to do something, as in (3c), typically expressed by a *to*-infinitive clause. Of these options, clearly (2b), memorizing something, is the most commonly used, prototypical one.

Table 1 shows the frequencies of these meanings in different World Englishes. It builds upon the components of the *International Corpus of English* (ICE) and documents absolute token frequencies (n), normalized frequencies per million words (which takes different corpus sizes into account; pmw)[5] and the relative proportion of the different meanings in each variety.

---

[5]  Calculating the wpm frequencies is actually not as straightforward a task as one might think, since for most corpora no 'official' figures identifying the corpus sizes are available. The target size for all ICE corpora is one million

Table 1. *Frequency distribution of the main meanings of* learn *in different varieties*

|  | 'receive info' | | | 'store info' | | | 'acquire ability' | | |
|---|---|---|---|---|---|---|---|---|---|
|  | n | pmw | % | n | pmw | % | n | pmw | % |
| GB | 31 | 29.9 | 10.7 | 224 | 216.1 | 77.5 | 34 | 32.8 | 11.8 |
| India | 24 | 21.8 | 6.3 | 324 | 295.1 | 85.0 | 33 | 30.1 | 8.7 |
| Hong Kong | 54 | 40.4 | 8.1 | 560 | 418.9 | 83.7 | 55 | 41.1 | 8.2 |
| Singapore | 27 | 24.7 | 5.6 | 376 | 343.5 | 78.0 | 79 | 72.2 | 16.4 |
| East Africa | 61 | 43.3 | 7.8 | 659 | 468.3 | 83.9 | 65 | 46.2 | 8.3 |
| Nigeria | 46 | 45.5 | 7.6 | 489 | 484.0 | 81.0 | 69 | 68.3 | 11.4 |

These figures show some interesting tendencies, instances of semasiological change (i.e. varying meanings attributed to the same form). Most importantly, in all New Englishes the core meaning 'store information' is used with higher proportions than in GB, and in the vast majority of instances the two more marginal meanings are less frequently used in the second-language-speaking varieties compared to GB. The only exception is one detail in Singapore, where the third meaning, 'acquire an ability', is apparently catching on more intensely than anywhere else, being used more commonly, in 16.4 percent of all tokens (as compared to between 8 and 11 percent elsewhere).

For Indian English (only) it is also possible to investigate change diachronically, since we have two megaword corpora from different time periods, the Kolhapur corpus from 1978 and the ICE corpus from about 2000.[6] Table 2 presents the pertinent figures. It confirms that in that period of twenty-plus years the core meaning 'store information' increased substantially while the two rarer meanings lost ground.

The results presented in tables 1 and 2 thus suggest two major possible trends of semasiological word meaning changes in the evolution of New Englishes. First, a 'focusing' process seems to be effective, so that in New Englishes as compared to the mother variety a core meaning (in this case 'store information') is gaining ground and used more widely, while less central meanings are losing ground and fall into disuse (and we may speculate whether in the long run they may be given up). Secondly, just like specific sounds, words or patterns get established in certain new varieties it is

words, but all corpora are somewhat larger for technical reasons (inclusion of metatext, utterances by non-locals, or tags); in particular, the Hong Kong (mainly because of inclusion of foreigners' talk in conversations; Kingsley Bolton, p.c.) and East Africa corpora are considerably larger. Furthermore, token counts also vary by software. I used the 'token used for word list' figure provided by WordSmith, and for ICE-Nigeria and ICE-East Africa figures given in the Manuals. For the record, the figures used are as follows: GB 1,036,649; HK 1,336,831; Singapore 1,094,614; India 1,097,785; Nigeria 1,010,382; East Africa 1,407,208.

[6] Both corpora are roughly equal in size (with a one-million-word target) but not quite equivalent stylistically (since Kolhapur is modeled upon the Brown corpus, with written texts only, while ICE has about 60 percent speech). I am interested in the relative proportions of meanings internally, and this seems largely unaffected by this detail and thus directly comparable.

Table 2. *Semantic change of* learn *in Indian English, 1978–2000*

|  | 'receive info' | | 'store info' | | 'acquire ability' | |
|---|---|---|---|---|---|---|
|  | n | % | n | % | n | % |
| 1978 Kolhapur | 28 | 10.9 | 206 | 79.8 | 24 | 9.3 |
| 2000 ICE | 24 | 6.3 | 324 | 85.0 | 33 | 8.7 |

possible that select word meanings get strengthened in specific varieties (only), so that in certain places certain meanings get more strongly associated with a word form (such as, here, the meaning 'acquire an ability' for *learn* in Singapore).

These results, tentative and limited as they are, show that clearly systematic changes in the evolution of word meanings in new varieties, general or locally specific focusing processes strengthening prototypical or also marginal associations, are to be expected (and deserve to be investigated further).

## 2.2 *Exemplary case studies II: semantic field diffusion*

A second potential type of change may involve shifting meaning relations between words in a 'semantic field'. This notion, also labeled differently, e.g. 'word field', 'lexical paradigm', etc., derives from a school of structural semantics (e.g. Lehrer 1974; Geckeler 1982; see Geeraerts 2009: ch. 2, Murphy 2010: 125–7) and assumes, adopting an onomasiological perspective (cf. Geeraerts 2009: 23–4 or the notion of an 'onomasiological profile' established in Grondelaers, Speelman & Geeraerts 2010: 1002), that meanings of words subdivide a field, a 'semantic space', amongst them, thus mutually defining their respective meanings and boundaries. For example, the field of 'dwellings' may be taken to encompass the lexemes *house*, *condo*, *hut*, *palace* and many more, and each word signals attributes which the others do not have. Earlier theory assumed that the meanings of such words delimit each other mutually rather precisely, while more recent thinking concedes that there are overlaps and semantic network relations (as described in WordNet, for example; Fellbaum 1998; cf. Lehrer 1974: 35; Schneider 1988a, I: 38–9).

For the present context of inquiring into possible semantic effects of variety emergence, a relevant research question will be: are there patterns of change in the constituency and mutual relations of sets of words within a word field when comparing donor and daughter varieties? Again, here are two examples, just illustrating options and scratching the surface.

My working assumption is that the verb pairs *recall*/*recollect* and *assume*/*suppose* are largely synonymous and thus subdivide the same semantic space amongst them in the manner of a semantic field, and the question then is whether there are noticeable changes in the frequency relations between these lexemes which show variety-specific

Table 3. *Absolute and relative frequencies of two synonymic verb pairs in different varieties*

|  | *recall* | *recollect* | *assume* | *suppose* |
|---|---|---|---|---|
| GB | 52 (94.5%) | 3 (5.5%) | 129 (26.2%) | 364 (73.8%) |
| India | 41 (60.3%) | 27 (39.7%) | 59 (10.4%) | 508 (89.6%) |
| Hong Kong | 107 (99.1%) | 1 (.09%) | 77 (22.8%) | 261 (77.2%) |
| Singapore | 63 (94.0%) | 4 (6.0%) | 123 (26.6%) | 339 (73.4%) |
| East Africa | 76 (97.4%) | 2 (2.6%) | 790 (30.4%) | 1,809 (69.6%) |
| Nigeria | 54 (96.4%) | 2 (3.6%) | 93 (35.1%) | 172 (64.9%) |

shifts in semantic spaces. Table 3 shows absolute and relative frequencies (based on the sum of tokens within each pair) for these two verb pairs for several varieties, again based on the ICE corpora.

Again, some interesting onomasiological shifts can be observed. In the first pair of semantically retrospective verbs, *recall* predominates strongly, and *recollect* is generally used rarely, at a rate of between about 1 and 5 percent of the time. There is one major exception, however: in India *recollect* shows a proportion of close to 40 percent of the time, suggesting that only there has this verb been regularly chosen to denote thinking back to the past, while elsewhere it is waning (strongly in Hong Kong, noticeably in Africa). In the second pair of verbs expressing insecure beliefs there are also changes, though less dramatic ones: in India, again, *suppose* (generally the preferred form) has been gaining ground even more strongly at the expense of *assume*, which seems recessive. In contrast, in the two African varieties (and especially in Nigeria) *assume* has been expanding and is used at higher proportions than elsewhere.

As before, the availability of the Kolhapur corpus allows a diachronic perspective for India, the variety which in both instances turned out to be particularly interesting. In this 1978 corpus the frequency of *recall* is 60 (83.3%), of *recollect* 12 (16.7%), and of *assume/ suppose* 87 (41.2%) and 124 (58.8%), respectively. Comparing these figures with the ICE-India ones in table 3 shows that in the late 1970s India's later preference for *suppose* had not yet started, while its predilection for *recollect* was well under way – and both processes have substantially gained momentum in the late twentieth century, with strongly increased proportions of both verbs in the ICE corpus.

Limited and exemplary as these case studies are, admittedly, they do show that there are systematic shifts of weights and emerging variety-specific preferences of word meanings as well, similar to what has generally been found for alternative syntactic choices or other linguistic options.

Thus, the present article will explore the question of lexicosemantic diffusion further. The focus will be on relationships within a semantic field, not on polysemy, and more specifically on the word field of 'prospective verbs'. In the light of this core distinction in semantic theorizing, this is thus an onomasiological study, asking how a relatively constant (or at least clearly circumscribed) set of meanings can be expressed by

varying forms (lexical choices), though, unavoidably, the study also entails related semasiological sidelines (asking which meanings are expressed by some form under varying circumstances).

## 3   Methodological procedure

### 3.1   Conceptual and methodological baseline: lexicosemantic variability

Lexical semantics, the study of word meanings, has been a core component of semantics (see, e.g., Riemer 2016) for decades, and several theoretical frameworks and methodological approaches have been developed (see Lehrer 1974 and Geeraerts 2009, amongst many others). In general, word meanings have been conceptualized as being fairly stable (except for change over long periods of time and the polysemy of many words).

An early study of lexicosemantic variability, including a variety perspective (differences between British and American English), is Schneider (1988a), which also has informed the present investigation to some extent. It adopts a feature-based componential analysis approach, assuming that word meanings are composites of minimal (not necessarily in an essentialist sense) meaning units, with varying degrees of overlaps of constituent features within and across words and with a strong emphasis on structural and extralinguistic context factors (verb complementation patterns, word co-occurrence classes and typical collocations), with which word meanings correlate quantitatively.[7] In many respects this approach corresponds to the one developed in quantitative cognitive semantics in the recent past (see Glynn 2010). For example, Glynn (2016) builds upon semantic feature analysis, 'introspection-based semantic categories (senses/nodes/types)' (417) and an observable usage profile of a lexeme (in his case the verb *annoy*), and develops correlations between these components – which is what I did then and do now as well.

Excellent research on lexicosemantic variability, mostly in Dutch, has been provided by the 'Quantitative Lexicology and Variational Linguistics' group at KU Leuven, formerly headed by Dirk Geeraerts, adopting a usage-based and cognitive framework and a corpus-based methodology. Geeraerts, Grondelaers & Bakema (1994) outline and apply a framework to analyze lexical variability, explicitly considering semasiological and onomasiological as well as formal and contextual variation, and emphasizing the importance of non-discreteness and prototypicality effects in all these instances. They also look into regional variability, comparing frequencies of components and other properties in select words in Netherlandic versus Belgian sources (1994: 105–15; cf. 146–53, 177–88). Similarly, Grondelaers, Speelman & Geeraerts (2010) outline a cognitively inspired 'sociolexicological' (988) theory of

---

[7]  However, this work, published in German, received scant attention internationally and was not promoted in circles of semanticists. Schneider (1988b) offers a summary discussion and illustration of the approach in English as applied to one lexeme.

lexical variability and change, which also highlights the importance of an onomasiological approach and the impact of metaphors and metonymy, prototypicality, radial sets and some types of semantic change. Some related and follow-up studies adopt a quantitative corpus-based approach by counting and correlating specific properties of a lexeme, such as semantic features and usage contexts, similar to the methodological procedure adopted in the present study (cf. Glynn & Fischer 2010; Glynn & Robinson 2014; Glynn 2016).

For the present study, an additional preliminary caveat is necessary: the notions of 'diffusion' and 'development' (in which we are interested) call for diachronic comparisons – but with very few exceptions in the study of World Englishes we have only synchronic corpus data available. Hence, what I can investigate are variability and differences but not really processes of change within a variety. The implicit working assumption, generally practiced in the discipline, has been that British English (or in very few cases American English) as the 'donor', origin variety represents the starting point, and differences between it and New Englishes represent effects of change in between. In the absence of reliable diachronically comparable corpora for most varieties in question (some of which are being compiled) this remains a defensible heuristic assumption – but we have to be aware of its limitations, and I will consider it when presenting results.

Theoretically, the present approach is ultimately inspired and framed by and feeds into the theory of complex dynamic systems (see, e.g., Kretzschmar 2015; Schneider 2020b, 2020c, and references mentioned there), which, together with the usage-based paradigm (Bybee 2010; Schmid 2020), I believe to be an adequate explanation of how language works and changes. For reasons of space, however, this connection will not be highlighted or discussed any further in this article.

### 3.2 Data and categorizations: investigating prospective verbs in World Englishes

Methodologically, for this particular study five types of components needed to be selected:

- a concisely defined set of semantically highly similar words, in this case verbs (a 'word field');
- a practically useful categorization of the semantic space in question;
- a clearly delimited categorization of relevant contextual factors, in this case defined as syntactic complementation patterns;
- a set of varieties for investigation; and
- suitable data bases (electronic corpora representing these varieties).

Inspired and informed by the earlier study mentioned (Schneider 1988a), I decided, adopting an onomasiological perspective, to select a group of verbs for investigation which is clearly circumscribed, appropriate in size, and both semantically and syntactically clearly structured and suitable for such an analysis, namely 'prospective verbs', i.e. verbs expressing thoughts about possible future events and states. The

following verbs are included: *intend*; *expect* (limited to prospective tokens, i.e. not if referring to a belief on the present or past); *look forward to*; *plan*; *contemplate* (also only if prospective); *anticipate*; *envisage*; and *envision*. These are the main verbs covering the semantic space of thinking about the future. At the same time, it is to be conceded, as theorists of lexical semantics have stated, that the demarcation of semantic fields is generally fuzzy and not clear-cut, and trying to delimitate them accurately is futile (Lehrer 1974: 35; Schneider 1988a, I: 38–9; Geeraerts, Grondelaer & Bakema 1994: 76–89, 118–34). In the present study two verbs were excluded for pragmatic reasons, namely *mean* (since the form is very frequent but very rarely used as a prospective verb) and *propose* (which mostly denotes a speech act and hardly ever a mental process).

The choice of varieties and corpora follows standard practice in many studies of World Englishes in the recent past. I investigate two first-language metropolitan varieties, putative 'donors', namely British English (GB/BrE) and American English (US/AmE), and four established and distinct second-language World Englishes (all from Asia), namely India (Ind), Singapore (Sing), Hong Kong (HK) and the Philippines (Phil). As data I used components of the *International Corpus of English* (ICE) project, a research initiative inspired and initiated by Sidney Greenbaum (1996), which since the 1990s has pursued the goal of compiling equally built electronic text collections which are representative of different varieties of English for comparison. Individual ICE corpora have a size of about one million words each, with equal proportions of text samples from different styles and genres, 60 percent of which are transcripts of spoken texts (see www.ice-corpora.uzh.ch/en.html).[8]

Based on the above definitions (all verb forms but only prospective meanings were considered), after manual pruning the database consists of 2,141 verb tokens, split up in table 4 by lexemes and varieties. Obviously, there are variety-specific lexeme preferences ($\chi^2 = 208.72$ at 35 df; p < .001).

Table 4. *The database: lemmata by varieties (token frequencies)*

|              | GB  | HK  | Ind | Phil | Sing | US  | sum   |
|--------------|-----|-----|-----|------|------|-----|-------|
| *anticipate*   | 21  | 16  | 17  | 18   | 18   | 27  | **117**   |
| *contemplate*  | 10  | 9   | 9   | 7    | 8    | 3   | **46**    |
| *envisage*     | 16  | 9   | 12  | 3    | 3    |     | **43**    |
| *envision*     | 1   | 1   |     | 7    | 1    | 12  | **22**    |
| *expect*       | 82  | 128 | 77  | 96   | 93   | 40  | **516**   |
| *intend*       | 74  | 58  | 51  | 62   | 89   | 26  | **360**   |
| *look fwd to*  | 56  | 72  | 25  | 40   | 71   | 13  | **277**   |
| *plan*         | 70  | 164 | 156 | 161  | 126  | 83  | **760**   |
| sum          | **330** | **457** | **347** | **394**  | **409**  | **204** | **2,141** |

---

[8]  ICE-USA lacks a spoken component; this was replaced by the *Santa Barbara Corpus of Spoken American English* (Du Bois *et al.* 2000–).

### *3.3 Prospective meanings*

Word meanings are notoriously difficult to define and grasp, since they are mental constructs which are not directly and objectively observable (see Taylor 2017). I employ an eclectic approach to semantic theory, proposing that the semantic space of prospective thinking can be circumscribed comprehensively by distinguishing six prototypical meanings (based on Schneider 1988a, slightly adjusted during the data coding process). These are to be understood as context-associated variant meanings of polysemic words, labeled by semantic theory 'sememes' (see Schneider 1988a, and sources quoted there) or 'allosemes' (Geeraerts 2009: 93). For each of them, I provide a concise label (for efficient future reference), a short definition and two examples each from the database, respectively. See examples (4) to (9).

(4) 'action self': think about a future activity of one's own, possibly to be realized soon
    (a) we … **look forward to welcoming you** to Singapore someday. (ICE-SIN W1B-017)
    (b) I am **planning to leave** Saturday morning (ICE-IND S1A-014)

(5) 'proposition': think about some event or state which will come true fairly soon (not an action of one's own)
    (a) Luna could … **envisage his countrymen identifying** with his bruised feelings (ICE-PHIL W2A-011)
    (b) you **expect it is going to cure** all problem (ICE-HK S2A-050)

(6) 'realize': think about some nominal entity, abstract or concrete, which will be realized/happen/come true in the future
    (a) … **anticipating failure**. (ICE-HK S1B-043)
    (b) some of his compatriots were already **contemplating collaboration** (ICE-PHIL W2E-03)

(7) 'purpose': think about something that will exist or be realized for a specific purpose in the future
    (a) Now **this** we **intended for dry skin** (ICE-PHIL S1B-047)
    (b) four day surgery suites have been **planned to handle more patients** (ICE-SIN S2B-016)

(8) 'receive': the subject believes they will own or receive some object soon
    (a) I **expect a minimum of Rd. 8500** (ICE-IND W1B-024)
    (b) I am **looking forward to your pictures** (ICE-HK W1B-012)

(9) 'be here': the subject believes that some person or object will be 'here', at the subject's location (the deictic center), soon
    (a) I **expect him** tonight (ICE-HK S2A-017)
    (b) they must also be **looking forward to you** (ICE-IND S1B-045)

### *3.4 Data extraction and coding*

All tokens of relevant verb occurrences were extracted from the corpora (using AntConc to produce KWIC concordances), transferred to Excel spreadsheets, and then coded in Excel

for relevant analytical categories and context factors. Verb lemma, meaning and variety were coded as defined above.

Syntactic verb complementation was focused on and coded as the most important context factor correlating with lexemes and meanings.[9] These syntactic patterns can be subdivided in various ways, depending on one's desired level of granularity. I opted for an intermediate one, which offers a reasonable balance between being sufficiently informative and practicable enough. The following thirteen variants of syntactic complementation were distinguished (again offering a label, a short definition and a brief example from the corpus):

- intr (intransitive): *I … want to plan ahead* (HK)
- intrPrpP (intransitive + prepositional phrase): *to plan for IT in education* (Phil)
- NP (plain noun phrase as direct object): *legislators expect further changes* (HK)
- NPtoinf (noun phrase + *to*-infinitive): *… expected young and old to touch his feet* (Ind)
- NPothComp (noun phrase + other complement): *the system is intended … for screening* (Phil)
- Øthtfincl (finite object clause without *that*): *he expected the final approval … would …* (HK)
- thatfincl (finite object clause with *that*): *he anticipates that they'll be open* (US)
- bareinfcl (bare infinitive clause): *you're not looking forward to get …* (Phil)
- toinfcl (*to*-infinitive clause): *we intend to fully utilize …* (Ind)
- Vingcl (nonfinite clause with verbal *-ing*-form): *what do you intend doing* (Ind)
- whcl (*wh* clause, finite or nonfinite): *we will contemplate … who is going to be …* (HK)
- comptrans (complex transitive complementation, = object + object complement): *can envisage Britain without a monarchy* (GB)
- so (form *so*): *they had planned so* (Ind)

### 3.5   Data analysis: procedural steps

Quantitative analyses, consisting of contingency tables and some multifactorial modeling in R (R Core Team 2020), searched for correlations and significant interactions[10] between lexemes, meanings, complementation patterns and varieties.

Initially, stable structural correlations between lemma, meaning and complementation needed to be worked out. This was done using Hierarchical Configurational Frequency Analysis (HCFA), assessing two-way interactions and

---

[9]  Other possibly relevant context factors include collocates and semantic classes of co-occurring lexemes, but these apply only in some tokens and are therefore very impractical to handle.

[10]  I am aware of recent recommendations in statistics (Wasserstein & Lazar 2016; Amrhein, Greenland & McShane 2019; Wasserstein, Schirm & Lazar 2019) questioning the binary classification of 'statistical significance' versus 'non-significance' divided by an arbitrary threshold value (typically p<.05), since it has led to many misunderstandings and misinterpretations and tends to downplay weaker effects. Instead, it is recommended to interpret p-values carefully, to accept uncertainty, and to be thoughtful (relative to contexts), modest and open (to transparency and the role of expert judgment) in one's interpretations. Taking this into consideration, I maintain, employ and report the significance distinctions implemented into the modeling software but attempt to phrase my observations in line with these recommendations.

configurations of all three factors. Secondly, regional preferences were sought for, triangulating two different approaches at statistical modeling, HCFA and Conditional Inference Trees (ctrees).[11]

HCFA (for which I used a script in R; Gries 2004) is a suitable method of multifactorial, non-parametric modeling for precisely the kind of data I have here, nominal categorizations. It has been assessed as 'a simple and powerful technique, yet surprisingly uncommon' (Glynn 2014: 318), being not really fashionable today (cf. Hoffmann 2011). By its nature it is a technique which is exploratory in the sense that it is meant for disclosing relationships and correlations within a dataset rather than being used for hypothesis-testing. Essentially it can be understood as multiple chi-square calculations upon multidimensional contingency tables. While chi-square identifies significant row–column correlations anywhere in a contingency table (metaphorically speaking it tells us somewhere in the house there is a party going on), HCFA calculates the effects of all individual factor configurations (cells), i.e. it identifies cells with significantly higher or lower observed than expected frequencies (Gries 2008: 241–54; Hoffmann 2011: 24–6; in other words, it tells us in which room the party is taking place and how lively it is). Figure 1 shows an exemplary output on an HCFA. It yields (amongst other data) factors and factor configurations (here between variety, verb lemma and verb meaning), the directionality of an effect (is the observed frequency higher or lower than expected?), the conventional symbolic shortcut representations of p-values as */**/*** for $p < .05/.01/.001$, respectively, and factor strength ('coefficient of pronouncedness' Q; Gries 2008: 252). So figure 1 shows that the verb *plan* is used substantially more often than expected with the 'action self' meaning in the Philippines, and less often with the 'proposition' meaning in Singapore and the Philippines.

'Conditional Inference Trees' (Tagliamonte & Baayen 2012; Levshina 2015: 291–300), commonly known as ctrees and widely employed in recent years in World Englishes modeling, are robust, non-parametric tree-structure models of regression and classification, an alternative to multiple regression especially for high-order

| 1 | variety | Vlemma | Vmeaning | Obs-exp | Dec | Q |
|---|---------|--------|----------|---------|-----|---|
| 121 | Ph | plan | actslf | > | *** | 0.029 |
| 126 | Sg | plan | propos | < | *** | 0.012 |
| 127 | Ph | plan | propos | < | *** | 0.012 |

Figure 1. Sample output of HCFA

[11] Reviewers suggested alternative ways of analyzing the data set, such as Principal Components Analysis or building models per region and comparing them. Certainly there are always various conceivable approaches, each with advantages or disadvantages of its own. I find the procedure adopted here simple, elegant and insightful, and suitable for the data set and purpose at hand.

interactions, small sample sizes and many predictors. The algorithm makes recursive binary splits of the data set until there are no more variables categorized as 'statistically significant' (i.e. with p < .05). Its outcome is commonly visualized as trees with branches. It returns the p-values of every split, and by means of bins below the nodes it shows relative proportions of dependent variable configurations. This will be illustrated in section 5.2.

Finally, a qualitative analysis follows, showcasing and interpreting indicative and interesting examples from the corpus.

## 4   Results

### 4.1   Correlations

Clearly there are systematic correlations between the main factors of analysis: certain verbs express certain meanings more than others and prefer certain complementation patterns, and meanings go together with constructional preferences. Strictly speaking, therefore, these are not independent variables in the statistical analyses. However, this is a caveat that applies to most, if not all, quantitative linguistic data. In sociolinguistic investigations for instance, there are strong covariation relations between the parameters of social class, gender and style. This section thus works out underlying lexicosemantic and structural relations in the word field in English in general, disregarding regional differences between World Englishes for the time being.

#### 4.1.1   Binary correlations

In the next three tables I employ a consistent color-coding scheme (illustrated in the legend to table 5) for very strong binary interactions of cell frequencies as returned by HCFA: white figures signal a frequency higher than expected, at p < .001 in very dark cells, and at p < .01 in somewhat lighter ones; conversely, black figures behind light grey shading shows a frequency lower than expected, at p < .01 (light grey cells) and p < .001 (very pale cells).

Table 5.   *Interactions between lemma and meaning*

|            | action self | be here | realize | propos | purpose | receive | sum   |
|------------|------------:|--------:|--------:|-------:|--------:|--------:|------:|
| *anticipate* | 7         | 0       | 67      | 43     | 0       | 0       | 117   |
| *contempl*   | 17        | 0       | 26      | 3      | 0       | 0       | 46    |
| *envisage*   | 6         | 0       | 21      | 16     | 0       | 0       | 43    |
| *envision*   | 0         | 0       | 11      | 11     | 0       | 0       | 22    |
| *expect*     | 5         | 18      | 170     | 281    | 0       | 42      | 516   |
| *intend*     | 230       | 0       | 17      | 27     | 83      | 3       | 360   |
| *look fwd to* | 157      | 2       | 94      | 8      | 0       | 16      | 277   |
| *plan*       | 466       | 0       | 249     | 33     | 9       | 3       | 760   |
| sum          | 888       | 20      | 655     | 422    | 92      | 64      | 2,141 |

Legend:
higher, p < .001   higher, p < .01   lower, p < .01   lower, p < .001

Table 5 highlights the interaction between the semasiological (which meanings can be expressed by individual verbs?) and the onomasiological (which verbs can express specific meanings?) perspectives. Not surprisingly, it shows clear preferences, i.e. binary interaction cell frequencies where p-values are so low that the null hypothesis of these correlations being chance products can be rejected with some confidence. For example, *expect* mostly means the belief in the future truth of a proposition, and also in the future presence of a person or an object; *intend* typically relates to a future action by oneself or something with an explicit purpose; and *anticipate* avoids 'action self'. Future activities by oneself are mainly verbalized by *intend* and *plan*, and less so by *look forward to*; realizing a nominal entity is most strongly associated with *anticipate*; and so on.

Table 6 identifies similar tendencies as to which patterns verbs prefer. For example, *anticipate* mainly goes with noun phrase and *that*-clause objects, while *contemplate* prefers noun phrases only; both disprefer infinitive complements. *Plan* is used mainly intransitively, often with prepositional phrase complements, or with an infinitive, but not normally with clausal complements. Similar tendencies apply for all verbs.

Table 7 draws an abstract picture of the semantics–syntax interface, showing how meanings are mainly encoded by specific patterns, irrespective of the verb involved – but in its mapping from meaning to form (though here this means structure, not lemma) this can still be seen as adopting an onomasiological perspective. Future actions by oneself are typically expressed by *to*-infinitives or verbal *-ing* clauses (but not finite object clauses); realizing a proposition calls for object clauses; a purpose associated with an object noun tends to be expressed by an infinitive or some other complement; and realizing or receiving something or believing that somebody will come implies noun phrase objects.

This section and the three tables presented here have probed into the interface between lexical choices, semantics and syntax: verbs have typical meanings and structural preferences, and certain syntactic patterns encode meanings better than others. This is not really surprising – the results confirm intuitions and subconscious knowledge we might have when thinking about these interrelationships (which, however, we do not normally do when reflecting on our language behavior). The innovative value of these results is that they document these relationships accurately, and they show that in most cases they are tendencies – strong tendencies, perhaps, but there are always choices and encoding alternatives available.

### 4.1.2  *Ternary correlations: lemma, meaning and complementation*

While the above results highlight rather abstract interrelations, three-way configurations showcase much more specific patterns and expressive options. The HCFA returns 21 ternary configurations with p-values lower than .001 (marked by ***), i.e. which occur exceptionally more often than expected, presented in table 8. A closer look at the configurations shows that the statistical procedure has identified what may be called prototypical constructions: characteristic interactions and typical ways of expressing prospective concepts which as competent speakers we are all familiar with and use

Table 6.  *Interactions between lemma and complementation*

| | intr | intrPrpP | NP | NPothComp | NPtoinfcl | Øthtfincl | thatfincl | infcl | Vingcl | whcl | comptrans | *so* | sum |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *anticipate* | 8 | 0 | 69 | 0 | 4 | 5 | 22 | 1 | 5 | 3 | 0 | 0 | 117 |
| *contemplate* | 2 | 1 | 33 | 0 | 0 | 0 | 0 | 0 | 9 | 1 | 0 | 0 | 46 |
| *envisage* | 2 | 0 | 21 | 4 | 0 | 0 | 7 | 2 | 5 | 0 | 2 | 0 | 43 |
| *envision* | 1 | 0 | 11 | 2 | 2 | 0 | 1 | 0 | 1 | 0 | 4 | 0 | 22 |
| *expect* | 4 | 0 | 233 | 5 | 226 | 39 | 5 | 3 | 0 | 0 | 1 | 0 | 516 |
| *intend* | 1 | 2 | 19 | 32 | 71 | 0 | 5 | 221 | 3 | 0 | 6 | 0 | 360 |
| *look fwd to* | 0 | 0 | 117 | 4 | 1 | 0 | 0 | 16 | 139 | 0 | 0 | 0 | 277 |
| *plan* | 42 | 52 | 219 | 6 | 18 | 1 | 2 | 402 | 5 | 10 | 2 | 1 | 760 |
| sum | 60 | 55 | 722 | 53 | 322 | 45 | 42 | 645 | 167 | 14 | 15 | 1 | 2,141 |

Table 7. *Interactions between meanings and complementation*

| | intr | intr PrpP | NP | NPothComp | infcl | NPto infcl | Øtht fincl | that fincl | Vingcl | whcl | comptrans | *so* | sum |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| action self | 18 | 25 | 19 | 0 | 643 | 3 | 4 | 4 | 164 | 6 | 2 | 0 | 888 |
| be here | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 20 |
| propos. | 26 | 2 | 12 | 15 | 0 | 268 | 41 | 38 | 3 | 8 | 8 | 1 | 422 |
| purpose | 0 | 2 | 0 | 33 | 2 | 50 | 0 | 0 | 0 | 0 | 5 | 0 | 92 |
| realize | 16 | 23 | 613 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 655 |
| receive | 0 | 3 | 58 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 64 |
| sum | 60 | 55 | 722 | 53 | 645 | 322 | 45 | 42 | 167 | 14 | 15 | 1 | 2,141 |

Table 8. *Very strong three-way interactions – prototypical constructions*

| Vlemma | Vmeaning | Vcomp | Obs-exp | Dec | Q | Example | Source |
|--------|----------|-------|---------|-----|-----|---------|--------|
| plan | actslf | infcl | > | *** | 0.15 | *I am planning to come back home* | Ind W1B-010 |
| expct | propos | NPtoinfcl | > | *** | 0.099 | *I ... expect the economy of Hong Kong to achieve real growth* | HK S1B-054 |
| intnd | actslf | infcl | > | *** | 0.083 | *where you intend to put the plants* | Sing W2D-016 |
| plan | realize | NP | > | *** | 0.064 | *we'll be able to plan a good program for ESA* | Phil S1A-086 |
| lkfwt | actslf | Vingcl | > | *** | 0.061 | *We ... look forward to working closely with all parties* | Sing W1B-016 |
| expct | realize | NP | > | *** | 0.055 | *legislators expect further changes* | HK S2B-002 |
| lkfwt | realize | NP | > | *** | 0.031 | *I'm looking forward to the good fight* | Sing S1B-044 |
| antic | realize | NP | > | *** | 0.025 | *He had not anticipated such a violent reaction* | Ind W2F-015 |
| intnd | purpos | NPtoinfcl | > | *** | 0.021 | *This ... compilation is intended to help the foreign visitor* | Sing S2B-042 |
| expct | receiv | NP | > | *** | 0.017 | *you don't expect anything from them* | Phil S1A-006 |
| expct | propos | Øthtfincl | > | *** | 0.016 | *you expect it is going to cure all problem* | HK S2A-050 |
| intnd | purpos | NPothComp | > | *** | 0.013 | *Large percentage of the river water is intended for irrigation of lakhs* | Ind W2A-040 |
| antic | propos | thatfincl | > | *** | 0.01 | *Gandhi ... anticipated that Mr Patil might defy the ... command* | Ind W2A-003 |
| cntpl | realize | NP | > | *** | 0.01 | *She's contemplating baptism* | Sing W1B-014 |
| expct | behere | NP | > | *** | 0.008 | *They are expecting around twelve people* | Sing S1A-024 |
| envag | realize | NP | > | *** | 0.008 | *Plans for the new airport envisage a ten-fold increase* | HK S2B-050 |
| plan | realize | intrPrpP | > | *** | 0.008 | *we should plan for a strong industrial recovery* | Ind S1B-054 |
| lkfwt | receiv | NP | > | *** | 0.006 | *I look forward to your continuous support* | HK W1B-022 |
| envag | propos | thatfincl | > | *** | 0.003 | *It can be envisaged that higher HIP pressure will bring about ...* | Sing W2A-040 |
| envon | propos | comptrans | > | *** | 0.002 | *The minister envisioned the museum precinct to be a lively part* | Sing S1B-049 |
| intnd | purpos | comptrans | > | *** | 0.002 | *The figures ... are ... simply intended as a guide* | GB W2B-040 |

regularly to encode typical relationships. For example, thinking about one's own possible future activities works well with *plan* or *intend* and an infinitive or with *look forward to* and a verbal *-ing* form; *expect* with a noun phrase typically means that something will be received or somebody will be here; something purpose-oriented in the future is preferably encoded by *intend* followed by a noun phrase and either an infinitive, an object complement in a complex transitive construction, or another complement; *contemplate* commonly implies a following noun phrase to come true; etc.

In sum, the correlations analysis has documented a range of systematic interactions between the three major factors. Verb lemmata express certain meanings predominantly, others less commonly; they require or allow specific verb complementation patterns predominantly, others less commonly; and verb meanings tend to be expressed by and go together with specific verb complementation patterns, and less commonly by/with others. In prototypical schematic constructions, of which the HCFA analysis has returned 21 patterns, specific verbs express specific meanings in specific complementation constructions.

## 5   Regional preferences

### 5.1   *Statistical modeling of regional preferences I: HCFA*

The goal of this section is to identify configurations in which variety makes a difference in interactions with two or three core factors. This requires a methodological caveat. Figure 2 shows an HCFA output for interactions of one lexeme (*expect*), one meaning ('proposition'), and several varieties. All interactions are very strong, with p mostly smaller than .001, so this might be taken to showcase regional differences. On closer inspection, however, it becomes clear that the strength of this effect is caused by the strong interaction between lexeme and meaning (in the first line), so that all subsequent interactions, with varieties added or not, fall out strongly significantly as well. However, the role of the varieties in these instances is coincidental, not causal, since the observed effect strength of these patterns is caused by the lemma–meaning configuration, not by the variety. In order to test the impact of variety as a causal factor, I therefore identify only configurations where variety makes a difference as compared

| 1 | variety | | Vlemma | | Vmeaning | | Obs-exp | | Dec | | Q | |
|---|---------|---|--------|---|----------|---|---------|---|-----|---|------|---|
| 36 | . | | expct | | propos | | > | | *** | | 0.088 | |
| 15 | HK | | expct | | propos | | > | | *** | | 0.022 | |
| 20 | GB | | expct | | propos | | > | | *** | | 0.017 | |
| 24 | Ph | | expct | | propos | | > | | *** | | 0.015 | |
| 34 | Sg | | expct | | propos | | > | | *** | | 0.012 | |
| 36 | In | | expct | | propos | | > | | *** | | 0.011 | |
| 648 | US | | expct | | propos | | > | | ** | | 0.008 | |

Figure 2. HCFA output for the interaction of *expect*, 'proposition' and several varieties

to the same configuration without variety: interactions with p< .05 involving variety were compared with the same configurations without variety, and only if the former show an effect which is stronger than the same configuration without variety (measured by comparing Q, the effect strength) or one which operates in a different direction altogether (represented by '<' vs. '>' in 'Obs-exp') is this interpreted as a configuration caused by variety.

Table 9 reports all relevant results (configurations with region different or stronger than without) for single factors. There are not many, none at all for meaning only, and the majority relate to metropolitan varieties. As to lexemes, *plan* is dispreferred in GB while *envision* and *anticipate* occur with higher than expected frequency in the US. The only lexical relationship involving one of the New Englishes concerns India, where *look forward to* is used less commonly than expected. For verb complementation distributions, three patterns have been selected. In the Philippines, complementation with V*ing* nonfinite clauses and with NPs + complements other than *to*-infinitives (generally dispreferred) are used more frequently than expected. In GB V*ing* complementation is also preferred.

Table 10 shows the impact of the factor variety for two-way configurations. All effects singled out by the statistical machinery are weak, and occur in metropolitan varieties only. *Envisage* relating to a proposition is weakly preferred in GB only; conversely, *envision* with a proposition and with the 'realize' meaning are preferred in the US only; this confirms the lexicographic description that *envisage* and *envision* are considered GB–US equivalents, respectively. The lemma *intend* with complex transitive complementation is weakly associated with GB. And as to meaning–complementation interactions, the 'action self' meaning is preferably expressed by means of an intransitive verb followed by a prepositional phrase in the US (e.g. *If you plan on taking short hikes …*), and in GB the 'purpose' meaning is associated with a complex transitive verb complementation more strongly than generally.

Table 9. *Modeling regional impact in HCFA: single factors*

| Factor and value | Configuration without region | | | Configuration with region | | | |
|---|---|---|---|---|---|---|---|
| | dir. | Dec | Q | variety | dir. | Dec | Q |
| Lemma | | | | | | | |
| *plan* | > | *** | 0.263 | GB | < | *** | 0.023 |
| *envision* | < | *** | 0.131 | US | > | *** | 0.005 |
| *anticipate* | < | *** | 0.08 | US | > | ** | 0.007 |
| *look fwd to* | > | n.s. | 0.005 | Ind | < | * | 0.009 |
| V complementation | | | | | | | |
| Ving | < | n.s. | 0.006 | GB | > | * | 0.009 |
| Ving | < | n.s. | 0.006 | Phil | > | * | 0.009 |
| NPothComp | < | *** | 0.064 | Phil | > | * | 0.006 |

Table 10. *Modeling regional impact in HCFA: two-way configurations*

| Factor and value | | Configuration without region | | | Configuration with region | | | |
|---|---|---|---|---|---|---|---|---|
| | | dir. | p | Q | variety | dir. | p | Q |
| Lemma | Meaning | | | | | | | |
| *envisage* | propos. | > | >.05 | 0.004 | GB | > | <.05 | 0.003 |
| *envision* | propos. | > | >.05 | 0.003 | US | > | <.01 | 0.003 |
| *envision* | realize | > | >.05 | 0.002 | US | > | <.05 | 0.003 |
| Lemma | V comp. | | | | | | | |
| *intend* | comp.trans. | > | >.05 | 0.002 | GB | > | <.05 | 0.002 |
| meaning | V comp. | | | | | | | |
| action self | intr PrepP | > | >.05 | 0.001 | US | > | <.05 | 0.004 |
| purpose | comp.trans. | > | <.05 | 0.002 | GB | > | <.01 | 0.002 |

The HCFA results for three-way interactions involving variety impact are similar: only three very weak effects were returned, and they are all in metropolitan varieties only. The configuration of *intend* with 'purpose' meaning and complex transitive complementation is weakly dispreferred in general but not so in GB. In the US, *envision* meaning 'realize' complemented by a NP and *plan* meaning 'action self' used intransitively with a PrepP are less common than elsewhere.

Overall, the HCFA analysis returns a few regional configurational preferences, but their number and intensity are relatively small. The investigation of single-factor effects yields a small set of lexical preferences (*envision*, *anticipate* in US) and dispreferences ( *plan* in GB, *look forward to* in India) and a small number of structural associations. For two-way and three-way configurations, the set of results is equally small and restricted to GB and the US only.

Despite the weakness of these effects, two results deserve to be noted. First, there are regional preferences involving Asian Englishes, but only for single factors: Indian English uses *look forward to* less commonly than other varieties, and Philippine English prefers complementation with V*ing* and NP+other complements. Secondly, it is noteworthy that two-way or three-way interactions are found in metropolitan donor varieties only, not at all in New Englishes. It may be speculated that this is a non-coincidental, persistent difference between the two variety types, since the older 'reference' varieties have had more time to stabilize, while New Englishes show signs of still being evolving, with less strongly entrenched and conventionalized habits and associations (see Schmid 2020 for an outline of the theoretical background of these notions).

## 5.2  *Statistical modeling of regional preferences II: ctrees*

Various ctrees have been produced, for variety and verb complementation as well as for variety, meaning and verb complementation as independent factors. However, with a

larger number of factors involved ctrees tend to get very complex and visually opaque, so I chose the most meaningful one for reproduction here: figure 3 presents a Conditional Inference Trees modeling of the data set for lemma as the dependent and meaning and variety as independent factors. This representation models a meaningful linguistic question: which lexeme is to be chosen to express a certain meaning in a given variety? Results as showcased by other representations are similar. Variety is consistently selected as significant (in the ctrees algorithm this is defined as $p < .05$) but always at lower-level splits; the impact of meaning and also verb complementation is obviously stronger.

The ctree shows that the primary splits on all levels are by meaning, which, not surprisingly, is thus selected as the strongest determinant of lexeme choice. Variety accounts for lower-level splits, but remarkably it turns out to be relevant for all meanings and meaning configurations except 'purpose'.

It is noteworthy that ctrees not only suggest categorizations based on internal similarities through the splits but also allow a qualitative association, showing with which form configurations a given parameter grouping (node) is associated. This is provided by the bins below the nodes, which show relative proportions of dependent variables associated with each cluster. Each of these bins (boxes) presents a histogram where the height of the bars identifies the relative frequency (as proportions of the sum total) of the verbs (as dependent variables), from left to right with the sequence *anticipate*, *contemplate*, *envisage*, *envision*, *expect*, *intend*, *look forward to* and *plan* in each bin. In other words, the characteristic choices or non-choices (frequency constellations) of the eight lemmata, represented by vertical bars, preferred by each configuration of parameters (independent variables) in the superordinate node are showcased.

Figure 4, a close-up of a section of figure 3, illustrates how the bins and their relationship to the independent variables are to be interpreted. It shows node 6, representations of the meaning 'receive', as determined variably by region (with p displayed) in its dependent bins node 7 and 8: Hong Kong is singled out because there this meaning is expressed mainly by *look forward to* (the seventh bar from the left in node 7) and also *expect* (i.e. the fifth bar from the left). In contrast, in all other varieties mainly *expect* and also *look forward to*, *intend*, and *plan* occur in this function (node 8, bars 5, 7, 6 and 8 from the left, respectively).

Further detailed correlations and preference relationships for significant cluster configurations (i.e. meanings preferably expressed by specific lexeme constellations in specific varieties) include the following examples:

- The meanings 'be here' and 'proposition', which jointly form node 3, show a split between the US, where they are mainly encoded by *expect* and also (much less strongly) *anticipate*, *envision* and *plan* (node 5) versus all others, where almost only *expect* occurs (node 4).
- The meaning 'realize' shows a first-level split (node 12) between the US (node 18, with *plan* and *anticipate* as main realizations) and all others; then, the Philippines (with
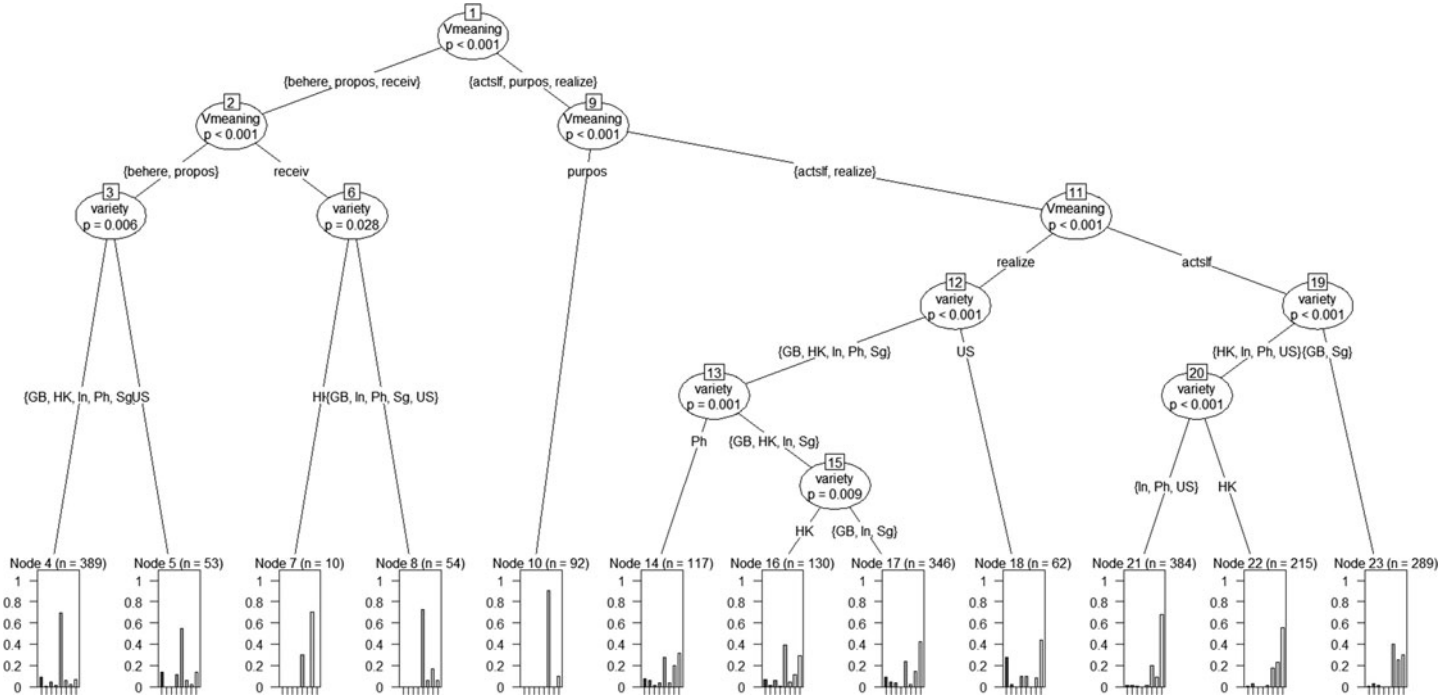
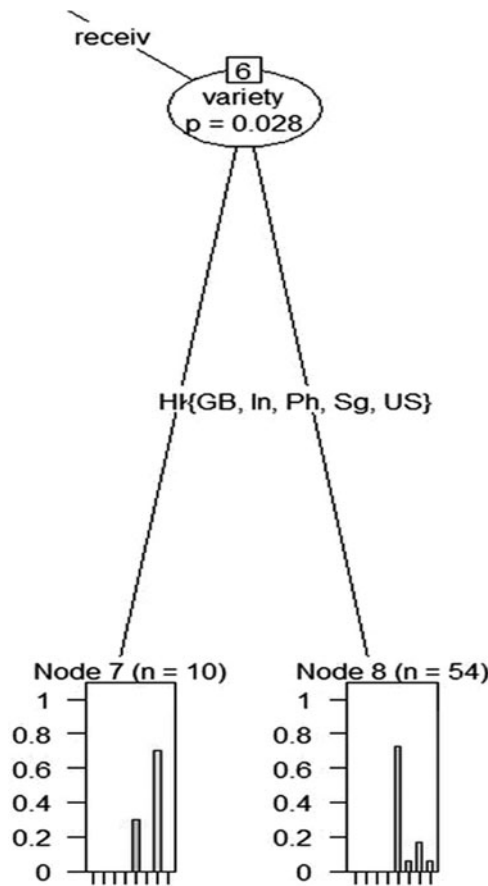Figure 3. Conditional Inference Tree (ctree) for lemma as dependent variable

Figure 4. Nodes 6, 7 and 8 (close-up from figure 3)

mainly *expect*, *look forward to* and *plan*, node 14) set themselves off (node 13), and finally (node 15), there is a difference between Hong Kong (node 16: mainly *expect* and *plan*) and GB, India and Singapore (where mainly *plan*, also *expect*, and less strongly some others are used; node 17).

- The meaning 'action self' first (node 19) sets off GB and Singapore as a cluster (with *intend*, *look forward to* and *plan* as its main manifestations, node 23); and amongst the others there is a division (node 20) between India, the Philippines and the US together (with mostly *plan* and some *intend* in this role, node 21) and Hong Kong (where *plan* and, less so, *look forward to* and *intend* predominate; node 22).

Thus, overall it is possible (if difficult) to interpret the ctree meaningfully: configurations of lemmata within clusters are complex but conditioned by specific hierarchies of independent factors. Ctrees modeling confirms the secondary but influential role of variety: individual varieties develop distinct but internally complex

configurations, similar to what is found in syntax (where variable frequency preferences have been documented for many different phenomena), in line with what the theory of Complex Dynamic Systems predicts (Kretzschmar 2015; Schneider 2020b, 2020c).

## 6   Qualitative analyses: regional patterns

Sophisticated statistical modeling is important because it identifies subtle effects and correlations which otherwise would not be visible. However, below the level of statistical significance and quantification there are also many linguistically interesting facts and observations which are worth pointing out and interpreting because they are structurally interesting and indicative, and possible embryonic indicators of possible future change and stabilization. Hence, in this section a few patterns which are noteworthy and possibly innovative but occur only rarely will be showcased.

The most interesting one of these, both structurally and because it occurs somewhat frequently and only in a specific variety type, is the sequence of *look forward to* and a plain infinitive, as exemplified in (10a–c).

(10)  (a)  I'm **looking forward to meet** you (HK W1B-014)
    (b)  We always **look forward to be** of service to you. (Sing W1B-028)
    (c)  The … Chief Minister … could now **look forward to carry** on his alliance government (Ind W2E-009)

Remarkably, this construction occurs in Asian Englishes only, and not at all in GB or the US. The database has 14 examples: 6 from Hong Kong (in 4 different texts), 5 from Singapore, 2 from India, and one from the Philippines. Five examples appear in spoken texts and 9 in writing – given the fact that the proportion of speech in ICE corpora is stronger (60 percent), this distribution implies that the construction appears to be stylistically neutral to formal, certainly not typically informal usage. A putative cause may be an obvious structural analogy, or restructuring: *to* in *look forward to* appears to be understood and reinterpreted as an infinitive marker (rather than a preposition). The number and regional as well as textual/stylistic spread of this pattern seems to be too frequent and regular to be disregarded as merely idiosyncrasies or 'errors', although the number of attestations is insufficient for any deeper analysis or substantial statement. Still, the regional distribution, the fact that this structure is found in second-language New Englishes only, suggests that here we might be identifying an embryonic structure, an innovative syntactic tendency which may be spreading in these varieties in the future. Time will tell – but it should be interesting to keep observing this development.

The other qualitative regional patterns to be reported here are idiosyncrasies, what may be termed 'structural hapax legomena': innovative patterns which occur only once (and in one case twice). These structures, illustrated in (11) to (14), are possibly interesting, but their interpretation has to remain speculative by necessity, given the sparsity of the evidence: are these just 'errors' – or can they also be regarded as embryonic traces of some innovative tendency? What is remarkable is that these constructions occur only in Asian Englishes, and are not found at all in GB or the US.

(11)  *expect* + NP bare infinitive clause (two examples):
    (a)   in a few minutes we could **expect the president come** out of the left side of the Senate (Phil S2A-005)
    (b)   The boy would not **expect the foreigner come** to his home (HK W1B-001)

(12)  Varying patterns with *plan*:
    (a)   *to* + Ving clause:
        but to resort to reducing the subsidies and to **plan to generating** funds on your own. (Ind S1B-040)
    (b)   clausal substitute (Halliday & Hasan 1976: 130–41):
        I think so. And uh they had **planned so** uhn. (Ind S1A-052)
    (c)   conditional clause complement, meaning 'action self':
        I'm **planning if** I just want to gain the experience of interview (HK S1A-092)
    (d)   mediopassive reading:
        Later we will see the uh Kwai Chung container port and uh the extension uh one more will be built in Lan uh Stonecutter Island and the **others are planning** in uh Tsing Yi Island and also Lantau Island (HK S2A-025)

(13)  Varying patterns with *intend*:
    (a)   **What I intend to is** that I just go through … (Sing S2A-041)
    (b)   He also wrote church **opera**. They are partly worship and partly operatic **that intend to performed** in church. (HK W1A-015)

(14)  Complementation of *contemplate*:
    Kumu's home, and … **contemplating MA** (Applied Ling.) (Sing W1B-003)

With the two examples of *expect* in (11), the infinitive marker *to* is omitted in nonfinite object clauses; a bare infinitive is used as the formal representation of the predicate. As shown in (12), the verb *plan* invites a variety of unusual complementation patterns. In (12a), instead of an infinitive a verbal *-ing* form follows *plan to*, so this seems a reversal of the pattern in (10). The example seems caused by a priming effect, since the preceding verb, *resort*, is also complemented by *to* plus a V*ing* form – but in the case of *resort* this conforms to established linguistic conventions (*resort* is standardly complemented by the preposition *to* plus a noun phrase, here realized by a verbal noun form), which is not the case for *plan* (which may take a *to*-infinitive but not *to* as a preposition). The same cause, a priming effect where an immediately preceding form is mentally still activated and thus simply copied, explains (12b): the verb *think* allows the word *so* as a clausal substitute form, replacing an object clause (Halliday & Hasan 1976: 130–41), but in mainstream English this is not established after *plan*. Furthermore, (12c) and (12d) also exemplify uses of *plan* which in standard English would not be accepted. There are several conventionalized options to express a future action of one's own after *plan* (see tables 6 and 7), but a finite *if*-conditional clause is not amongst them. And *plan* is also not one of the 'mediopassive' verbs in English which encode a passive reading by an active form, as in (12d), where '(ports) *are planning*' is to be understood as 'are being planned'. Examples (13a) and (13b) display

incoherent structures, apparently a confusion of complementation patterns, involving the verb *intend*. In (13a), *intend* is followed by *to* but not by an infinitive, possibly to be understood as a NP complement of *to* or as an object of an understood but deleted verb (*what I intend to do is…*). In (13b), active and passive forms are confused, with *be* lacking. And finally, (14) is semantically incoherent: *contemplate* needs to be complemented by an activity or process noun, not by a noun denoting the result of an activity: *contemplate working/studying for an MA* would be fine, *contemplating MA* sounds incomplete.

As stated earlier, it is difficult to assess these patterns convincingly. While the pattern illustrated in (10) seems on its ways towards being established as a new, firm habit in some New Englishes, the ones in (13) and possibly the ones in (12) and also (11) seem products of creative usage and limited exposure to standard conventions. In any case, it may be rewarding to watch out for further occurrences of (some of) these constructions.

## 7   Summary and conclusion

This article has opened a new research branch in the investigation of World Englishes. It has shown that word meanings and their associated contextual patterns, especially verb complementation structures, also vary and change in systematic ways, and also across varieties in the process of linguistic transmission from donor varieties to New Englishes. Similar to what has been found frequently for syntactic variability, there are persistent frequency differences and varying preference tendencies in the use of lexemes, their meanings, and their associated patterns of structural behavior.

Systematic relations between lemmata, meanings and complementation options are strong and stable – an observation which describes the relation between these factors in English in general. By comparison, regional differences are much weaker (not surprisingly, since the diffusion process is subtle and operates below the level of linguistic awareness), but they do exist and persist, and have been shown to be effective in various ways. Statistical techniques have singled out some of the patterns described as highly significant – which suggests that they may be emergent schematic constructions in Asian Englishes. For example, the HCFA has documented a dispreference for *look forward to* in Indian English and preferences for verbal *-ing* complements and NPs with complements other than *to*-infinitives in Philippine English. Modeling with ctrees has identified several (fuzzy) clusters of dependent variable groupings which are strongly associated with specific independent variable configurations, and in general it has worked out that variety is a relevant factor, even if it is secondary in strength to meaning. Finally, qualitative documentation has identified a range of idiosyncratic patterns which may simply be viewed (and largely disregarded) as instances of spontaneous creation and deviation, but clearly they may also be emergent patterns, early embryonic indicators of what may become future regularities in some varieties. The strongest case in point along these lines, a pattern for which there is sufficiently strong evidence in several varieties and styles, is the construction *look forward to* plus infinitive in Asian Englishes, which appears to be on its way to becoming established more regularly there.

Overall, no clear overarching regional patterns but some tendencies, perhaps emerging ones, are discernible. Some preferences manifest themselves in the older, metropolitan varieties only, and in some observations and nodes US English in particular sets itself off from other varieties. India appears to stand out in some ways, e.g. in adopting structural innovations; and less so this can also be observed for Hong Kong and the Philippines, and least for Singapore (which often sides with British usage).

Analyzing meanings (of words and constructions) is difficult since they manifest themselves in speakers' minds and are not directly and objectively accessible and observable (unlike sounds or syntactic forms and structures), so semantic analysis requires some sort of definition of entities and analytical procedures which allows linguists to categorize them and handle them as units for analysis. But of course meanings are an essential part of human language; they are what we wish and need to encode using sounds, words and constructions. In line with researchers in cognitive variationist lexicology I argue, therefore, that they deserve and have to be a core component of investigations into language variation and change and linguistic diffusion, including the processes that have produced new World Englishes. The topic of lexicosemantic variability and diffusion in variationist linguistics and in World Englishes is underresearched but promising, in need of further investigation.

*Author's address:*

*Department of English and American Studies*
*University of Regensburg*
*Universitätsstraße 31*
*93053 Regensburg*
*Germany*
*edgar.schneider@ur.de*

## References

Amrhein, Valentin, Sander Greenland & Blake McShane. 2019. Retire statistical significance. *Nature* 567, 305–7.

Bao, Zhiming. 2015. *The making of vernacular Singapore English: System, transfer and filter*. Cambridge: Cambridge University Press.

Blank, Andreas. 1999. Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change. In Andreas Blank & Peter Koch (eds.), *Historical semantics and cognition*, 61–90. Berlin: Mouton de Gruyter.

Bruckmaier, Elisabeth. 2017. *Getting at GET in World Englishes: A corpus-based semasiological-syntactic analysis*. Berlin: De Gruyter.

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Coleman, Linda & Paul Kay. 1981. Prototype semantics: The English word *lie*. *Language* 57, 26–44.

Du Bois, John W. 2000–. *Santa Barbara Corpus of Spoken American English*. www.linguistics.ucsb.edu/research/santa-barbara-corpus

Fellbaum, Christiane (ed.). 1998. *WordNet*. Cambridge, MA: MIT Press.

Geckeler, Horst. 1982. *Strukturelle Semantik und Wortfeldtheorie*. Munich: Fink.

Geeraerts, Dirk, Stefan Grondelaers & Peter Bakema. 1994. *The structure of lexical variation: Meaning, naming and context*. Berlin: Mouton de Gruyter.

Geeraerts, Dirk. 2009. *Theories of lexical semantics*. Oxford: Oxford University Press.

Glynn, Dylan. 2010. Corpus-driven cognitive semantics: Introduction to the field. In Glynn & Fischer (eds.), 1–42.

Glynn Dylan. 2014. Techniques and tools: Corpus methods and statistics for semantics. In Glynn & Robinson (eds.), 307–41.

Glynn, Dylan. 2016. Quantifying polysemy: Corpus methodology for prototype theory. *Folia Linguistica* 50, 413–47.

Glynn, Dylan, & Kerstin Fischer (eds.). 2010. *Quantitative methods in cognitive semantics: Corpus-driven methods*. Berlin: De Gruyter Mouton.

Glynn, Dylan & Justyna Robinson (eds.). 2014. *Corpus methods for semantics: Quantitative studies in polysemy and synonymy.* Amsterdam: John Benjamins.

Greenbaum, Sidney (ed.). 1996. *Comparing English worldwide: The International Corpus of English*. Oxford: Clarendon Press.

Gries, Stefan Th. 2004. *HCFA 3.2 – A program for hierarchical configurational frequency analysis for R for Windows*. Available from the author upon request.

Gries, Stefan Th. 2008. *Statistik für Sprachwissenschaftler*. Göttingen: Vandenhoeck & Ruprecht.

Grondelaers, Stefan, Dirk Speelman & Dirk Geeraerts. 2010. Lexical variation and change. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 988–1011. Oxford: Oxford University Press.

Halliday M. A. K. & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.

Hoffmann, Thomas. 2011. *Preposition placement in English: A usage-based approach*. Cambridge: Cambridge University Press.

Hoffmann, Thomas. 2021. *The cognitive foundation of post-colonial Englishes: Construction grammar as the cognitive theory for the dynamic model*. Cambridge: Cambridge University Process.

ICE: *International Corpus of English*. www.ice-corpora.uzh.ch/en.html (accessed 6 December 2021).

Kirkpatrick, Andy. 2007. *World Englishes: Implications for international communication and English language teaching*. Cambridge: Cambridge University Press.

Kirkpatrick, Andy (ed.). 2010. *The Routledge handbook of World Englishes*. London: Routledge.

Klemola, Juhani, Markku Filppula & Devyani Sharma (eds.). 2017. *The Oxford handbook of World Englishes*. Oxford: Oxford University Press.

Kretzschmar, William. 2015. *Language and complex systems*. Cambridge: Cambridge University Press.

Lehrer, Adrienne. 1974. *Semantic fields and lexical structure*. Amsterdam: North-Holland.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Mehl, Seth. 2018. Corpus onomasiology in world Englishes and the concrete verbs *make* and *give*. *World Englishes* 37, 185–206.

Mesthrie, Rajend & Rakesh Bhatt. 2008. *World Englishes*. Cambridge: Cambridge University Press.

Murphy, M. Lynne. 2010. *Lexical meaning*. Cambridge: Cambridge University Press.

Nelson, Cecil, Zoya Proshina & Daniel Davis (eds.). 2020. *The handbook of World Englishes*, 2nd edn. Malden, MA, and Oxford: Wiley-Blackwell.

R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. www.r-project.org (accessed 31 January 2020).

Riemer, Nick (ed.). 2016. *The Routledge handbook of semantics*. London and New York: Routledge.

Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system*. Oxford: Oxford University Press.

Schneider, Edgar W. 1988a. *Variabilität, Polysemie und Unschärfe der Wortbedeutung*, vol. I: *Theoretische und methodische Grundlagen*; vol. II: *Studien zur lexikalischen Semantik der mentalen Verben des Englischen*. Tübingen: Niemeyer.

Schneider, Edgar W. 1988b. On polysemy in English, considering *consider*. In Werner Hüllen & Rainer Schulze (eds.), *Understanding the lexicon: Meaning, sense and world knowledge in lexical semantics*, 157–69. Tübingen: Niemeyer.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.

Schneider, Edgar W. 2020a. *English around the world: An introduction*. Cambridge: Cambridge University Press.

Schneider, Edgar W. 2020b. *Calling* Englishes *as* Complex Dynamic Systems: Diffusion and restructuring. In Anna Mauranen & Svetlana Vetchinnikova (eds.), *Language change: The impact of English as a lingua franca*, 15–43. Cambridge: Cambridge University Press.

Schneider, Edgar W. 2020c. Meanderings from early English to World Englishes: A Complex Systems perspective on morphosyntactic changes in *wh*-pronouns. In Peter Grund & Megan E. Hartman (eds.), *Studies in the history of the English language VIII: Boundaries and boundary-crossings in the history of English*, 73–105. Berlin and New York: De Gruyter Mouton.

Schneider, Edgar W., Bernd Kortmann, Clive Upton, Rajend Mesthrie & Kate Burridge (eds.). 2004. *A handbook of varieties of English: A multimedia reference tool*, vol. I: *Phonology*; vol. II: *Morphology and syntax*. Berlin and New York: De Gruyter Mouton.

Schreier, Daniel, Marianne Hundt & Edgar W. Schneider (eds.). 2020. *The Cambridge handbook of World Englishes*. Cambridge: Cambridge University Press.

Stern, Gustaf. 1931. *Meaning and change of meaning, with special reference to the English language*. Bloomington: Indiana University Press.

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24, 135–78.

Taylor, John R. 2017. Lexical semantics. In Barbara Dancygier (ed.), *The Cambridge handbook of cognitive linguistics*, 246–61. Cambridge: Cambridge University Press.

Wasserman, Ronel. 2020. The historical development of South African English: Semantic features. In Raymond Hickey (ed.), *English in multilingual South Africa: The linguistics of contact and change*, 52–73. Cambridge: Cambridge University Press.

Wasserstein, Ronald & Nicole Lazar. 2016. The ASA's statement on p-values: Context, process, and purpose. *The American Statistician* 70, 129–33.

Wasserstein, Ronald, Allen Schirm & Nicole Lazar. 2019. Moving to a world beyond 'p < 0.05'. *The American Statistician* 73, sup1, 1–19.

Werner, Janina & Joybrato Mukherjee. 2012. Highly polysemous verbs in New Englishes: A corpus-based pilot study of Sri Lankan and Indian English. In Sebastian Hoffman (ed.), *English corpus linguistics: Looking back, moving forward*, 249–66. Amsterdam: Rodopi.