

# *A topic which I want to know more about* – preposition placement in finite WH-relative clauses in World Englishes<sup>1</sup>

VICTORIA MUßEMANN 

Catholic University of Eichstätt-Ingolstadt

(Received 17 July 2023; revised 11 December 2023)

The present article analyzes the use of preposition stranding (*the world which we live in*) and pied-piping (*the world in which we live*) in finite WH-relative clauses in twelve varieties of English. In the light of previous studies, it assumes that the strength of processing constraints and formality effects that drive speakers' constructional choices should correlate with Dynamic Model stages (Schneider 2007). However, drawing on data from the *International Corpus of English* (ICE) and using mixed-effects logistic regression analysis, the study shows that processing factors affect speakers of all Dynamic Model stages in a very similar way. At the same time, clear differences between variety stages are observed with respect to formality and topic, which strongly affect Phases IV and V but not Phase III. These results are interpreted from a Usage-based Construction Grammar perspective.

**Keywords:** preposition placement, Construction Grammar, World Englishes, Dynamic Model

## 1 Introduction

The term 'preposition placement' (Hoffmann 2011) refers to the structural alternation between preposition stranding and preposition pied-piping in English. In relative clauses a preposition can either remain in-situ without its complement (i.e. 'stranded' as in example (1a)), or it can be placed in front of the relative pronoun, which is referred to as 'pied-piping' (1b) (Hornstein & Weinberg 1981; Ross 1986; Pullum & Huddleston 2002; Hoffmann 2006, 2007, 2011, 2013).

- (1) (a) the real world **which** we all have to live **in** (ICE-GB:S1B-035 #67:1:C)
- (b) the social world **in which** we live (ICE-GB:S1B-028 #79:1:B)

Although preposition placement in English has been studied widely, it has not received much attention from a World Englishes perspective. The few existing studies (Suárez-Gómez 2014, 2015 on selected Southeast Asian Englishes; Dayag 2016 on Philippine English) analyzed individual varieties only and were not based on multifactorial statistical analyses. Jach (2018, 2021) focused specifically on learner

<sup>1</sup> The present study was funded by a German Research Foundation (DFG) grant (HO 3904/7-1).

Englishes. An exception is Hoffmann's (2011) investigation of preposition placement in British English (BrE) and Kenyan English (KenE). Adopting a Usage-based Construction Grammar approach, Hoffmann (2011) showed that differences between the BrE and KenE networks of preposition placement constructions result from stylistic effects as well as processing factors. In line with a number of studies that point to a lack of register distinctions in L2 Englishes (e.g. Gilquin & Paquot 2008; Xiao 2009; Van Rooy *et al.* 2010; Buregeya 2019: 171), he found that BrE speakers favor preposition pied-piping in formal but stranding in informal relative clauses, whereas KenE speakers show only a very weak stylistic effect and prefer the pied-piped construction regardless of the text type (Hoffmann 2011: 155, 167). Furthermore, KenE speakers rated non-prototypical constructional choices such as pied-piping with prepositional verbs and stranding with adjunct PPs, which involve an increase in processing efforts, lower than BrE speakers (Hoffmann 2011: 187). In another experimental study the same effect was observed for German L2 learners (Hoffmann 2013: 114–16). Hoffmann (2011: 270) thus concludes that L2 speakers, who typically receive less language input, 'tend to favour ... prototypical realizations of preposition placement more than the British speakers'. This claim also receives support from Jach's (2018: 286) acceptability study, in which L2 German and Chinese speakers rated pied-piping with prepositional verbs, i.e. constructions that are difficult to process, lower than L1 English speakers.

However, with regard to other phenomena of syntactic alternation, research on the effect of processing constraints on L1 and L2 speakers of English has not produced clear results. On the one hand, there are some studies (Wulff *et al.* 2014, 2018; Wulff & Gries 2019) that attribute differences in constructional preferences between L1 speakers and L2 learners of English at least in part to the stronger effect of processing factors on the L2 speakers. On the other hand, studies on the genitive (Szmrecsanyi *et al.* 2016; Heller *et al.* 2017) or the dative alternation (Bernaisch *et al.* 2014; Röthlisberger *et al.* 2017) did not find clear processing-driven differences between L1 and L2 Englishes but claim that 'factors determining processing and, thus, ultimately constructional choices are widely applicable to all varieties of English' (Bernaisch *et al.* 2014: 28). Furthermore, Dubois *et al.* (2023: 20) pointed out that studies on the role of processing factors in L2 language should take into account different proficiency levels of learners, which reflect differences in exposure to input.

Against the background of these studies, the present analysis will now go beyond a strict L1-L2 variety distinction and extend Hoffmann's (2011) approach to a wider range of L1 and L2 Englishes at different stages of Schneider's (2007) Dynamic Model. Drawing on data from the *International Corpus of English* (ICE; Greenbaum & Nelson 1996), it will explore the use of preposition stranding and pied-piping in finite WH-relative clauses (RCs) in twelve varieties of English. Adopting a Usage-based Construction Grammar approach, it aims to assess how processing constraints and stylistic factors affect the constructional choices of speakers of varieties of English worldwide.

## 2 Theoretical background

### 2.1 *Schneider's (2007) Dynamic Model*

This study follows Schneider's (2007) classification of varieties of English as outlined in his Dynamic Model of the evolution of postcolonial Englishes (PCEs). According to Schneider (2007), all PCEs undergo a uniform evolutionary process through up to five developmental stages ('foundation', 'exonormative stabilization', 'nativization', 'endonormative stabilization', 'differentiation') in which linguistic changes are strongly associated with sociopolitical parameters, identity reconstructions and sociolinguistic conditions that characterize the contact setting between settler (STL) and indigenous (IND) strands. The present study focuses on varieties in Phases III, IV and V of the Dynamic Model. In Phase III, the STL and IDG strand groups gradually start to adopt a shared identity. This also involves increased linguistic contact, which ultimately leads to 'structural nativization' (Schneider 2007: 44), i.e. the development of structural innovation 'at the interface between grammar and lexis' (Schneider 2007: 46). In Phase IV, local linguistic forms are increasingly accepted and associated with prestige. Finally, Phase V is characterized by the emergence of dialects as markers of group identities. The sociolinguistic conditions associated with each Dynamic Model phase imply that the domains of use of English and the English language input speakers are exposed to increase as varieties evolve along the evolutionary cycle. Table 1 gives an overview of the varieties investigated in the present study categorized according to Dynamic Model phases.

### 2.2 *Usage-based Construction Grammar*

According to Usage-based Construction Grammar (Croft 2001; Goldberg 2006; Diessel 2019; Hoffmann 2022), speakers store all linguistic knowledge in the form of constructions, which are defined as pairings of form and meaning (Croft & Cruse 2004: 255; Goldberg 2006: 5). Construction Grammar assumes a 'syntax–lexicon continuum' (Croft & Cruse 2004: 256) ranging from completely schematic constructions such as the resultative construction ([X V Y Z] – 'X causes Y to become

Table 1. *ICE varieties investigated in this study along Schneider's (2007: 113–250) Dynamic Model*

Phase III	Phase IV	Phase V
ICE-East Africa (Kenya + Tanzania)	ICE-Jamaica	ICE-Canada
ICE-Hong Kong	ICE-Singapore	ICE-Great Britain
ICE-India		ICE-Ireland
ICE-Nigeria		ICE-New Zealand
ICE-Philippines		

Z by V-ing'), whose slots can be filled freely, to the most substantive constructions, which are phonologically fully specified (e.g. apple [æpl] – 'apple'; Hoffmann & Trousdale 2013: 2). Drawing on language-independent cognitive principles such as categorization, chunking and rich memory storage, usage-based approaches assume that abstract mental representations of constructions can only emerge from bottom-up generalizations over more substantive constructions (Bybee 2010: 7–9). While high token frequency leads to the independent storage of substantive constructions, abstract constructional templates can become entrenched if a construction is encountered with a high type frequency (Croft & Cruse 2004: 292–3; Bybee 2010: 95–6).

### 2.3 Preposition placement from a Usage-based Construction Grammar perspective

Turning more specifically to preposition placement, Hoffmann (2011: 264–75) provides a detailed, usage-based account of the construction networks of BrE and KenE speakers. Importantly, the term preposition placement is a theoretical construct that does not exist as an abstract construction in the mental grammars of speakers (Hoffmann 2011: 264). Instead, stranded and pied-piped prepositions occur across a wide range of clause types many of which only permit either stranding or pied-piping (for an overview see e.g. Pullum & Huddleston 2002: 627–8).

With respect to finite relative clauses, WH-relatives allow for both variants, with pied-piping generally being considered the more formal option in Standard British and American English (Quirk *et al.* 1985: 664; Biber *et al.* 1999: 107). Hence, based on the results of his corpus study, Hoffmann (2011: 268) suggests that BrE speakers possess a schematic informal stranded RC construction as well as a schematic formal pied-piped RC construction. In addition, stranding being obligatory in *that*- (2) and zero-relatives (3) (Pullum & Huddleston 2002: 627), from a usage-based perspective, the construction network can also be said to contain a stranded *that*- and a stranded zero-relative clause construction.

- (2) (a) the world **that** I was working **in** (ICE-GB:S1A-001 #35:1:B)
- (b) \* the world **in that** I was working
- (3) (a) the disabled people **Ø** you were working **with** (ICE-GB:S1A-002 #80:1:A)
- (b) \* the disabled people **with Ø** you were working

Moreover, regarding locative and temporal relative constructions, speakers can alternatively use constructions in which the WH-pronoun and the preposition are replaced by a 'relative adverb' (Biber *et al.* 1999: 60), i.e. by *when* (4b) or *where* (5b) (Biber *et al.* 1999: 624; Hoffmann 2011: 37).

- (4) (a) the moment **at which** Art chose life (ICE-NZ:W2F-006#130:1)
- (b) a moment **when** her expression switches from sympathetic consideration to one of decisiveness (ICE-GB:W2F-019 #31:1)
- (5) (a) a place **in which** forty-nine percent of the people didn't want to be there (ICE-CAN: S2A-034#137:2:A)
- (b) the place **where** we already are (ICE-GB:S1A-082 #62:1:A)

In line with Diessel's (2019: 199) assumption that 'the grammar network involves horizontal, or lateral, relations between semantically or formally similar constructions at the same level of abstraction', all these constructions can be said to be interconnected by horizontal associations.

Furthermore, speakers store partly schematic constructions such as a pied-piped *in-which* RC construction (Jach 2021: 360), in which only the P + WH string is phonologically specified, or substantive antecedent + P + WH chunks such as *way in which* (6) or *extent to which* (7) (Hoffmann 2011: 160–5). While RCs are normally treated as a combination of two independent elements, namely the antecedent noun head (e.g. *way* in (6)) and the relative clause (e.g. *in which the actual process is going to run*), this indicates that speakers also store chunks that cut across such traditional syntactic categories.

(6) the **way in which** the actual process is going to run (ICE-GB:S1B-020 #115:1:A)

(7) the **extent to which** demand will level off or fall in the early 1990s (ICE-GB:W2A-015 #14:1)

However, Hoffmann (2011: 265–75) also shows that the construction networks of BrE and KenE speakers are not identical. For instance, the KenE construction network lacks an entrenched abstract stranded RC construction that is associated with informal contexts, which can be explained by the higher processing cost associated with stranded RC constructions (see section 2.4). This preference for more prototypical constructions in the L2 variety could be due to the fact that the construction networks of KenE speakers, who receive less input than L1 speakers, are not as deeply entrenched as those of BrE speakers (Hoffmann 2011: 275), which means that constructions cannot be activated as automatically and as easily (Langacker 2008: 16; Schmid 2020: 43, 213–14). Since the use of English in L2 varieties may thus involve higher cognitive cost, these speakers may have a stronger preference for prototypical constructions, i.e. for variants that involve a reduction in processing effort. Another factor that may contribute to this effect could be the formal instruction L2 speakers are typically exposed to, which may prime speakers of less advanced varieties towards more prototypical constructional choices.<sup>2</sup> Based on Hoffmann's (2011) findings, it can thus be expected that speakers of English worldwide differ from each other with regard to the entrenchment of stranded and pied-piped RC constructions due to the specific type and amount of input that is associated with different evolutionary stages of Schneider's model.

## 2.4 Preposition placement and processing factors

According to Hawkins' (2004: 3) 'Performance-Grammar Correspondence Hypothesis', language is strongly shaped by processing effects. Processing factors are also highly relevant in the choice of preposition placement constructions (Gries 2002; Hoffmann 2011: 93–8). RCs belong to the so-called 'filler-gap' constructions (Pollard & Sag

<sup>2</sup> I am grateful to an anonymous reviewer for this suggestion.

1994: 157), which are difficult to process. Not only has the ‘filler’ (e.g. *which* in (8a), *with which* in (8b)) to be matched with the corresponding ‘gap’, i.e. the position in which the element represented by the filler would be found in a declarative sentence, but simultaneously the material standing in the path from the filler to the gap also has to be processed (Hawkins 1999: 246–7). In these complex environments, preposition pied-piping offers a processing advantage over stranding because it avoids garden paths. As the human processor always aims to identify the earliest possible gap site (Hawkins 1999: 247), in the stranded example (8a) it is likely that the filler *which* is wrongly identified as the object of the main verb *win* (Hawkins 1999: 247, 277). In contrast, the pied-piped preposition in (8b) ensures that such a misanalysis is avoided (also see Hawkins’ (1999: 277) ‘Avoid Competing Subcategorizers’ principle; examples in (8) based on Hawkins 1999: 277).

- (8) (a) a new set of pipes [**which**]<sub>i</sub> he [wins (O<sub>i</sub>) a music competition **with** O<sub>i</sub>]  
 (b) a new set of pipes [**with which**]<sub>i</sub> he [wins a music competition O<sub>i</sub>]  
 (ICE-IRE:W2A-008\$A)

Compared to interrogative clauses, for example, RCs are particularly complex as they do not only require processing of the filler-gap domain (FGD) but also of the co-indexation domain of the antecedent noun phrase (NP) and the corresponding relativizer (i.e. of *a new set of pipes* and *which* in (8); Hawkins 2004: 199). This explains why RCs are prototypical pied-piping contexts (Trotta 2000: 55–7; Pullum & Huddleston 2002: 628–9; Hoffmann 2011: 155–6). Even greater processing efforts are required in restrictive RCs, in which the interpretation of the antecedent noun is dependent on the parsing of the RC (Hawkins 2004: 150). Consequently, restrictive RCs (9a) favor pied-piping even more than non-restrictive RCs (9b) (Hoffmann 2011: 169–70).

- (9) (a) people with whom I might have a more meaningful set of conversations (ICE-CAN: W1B-012#47:2)  
 (b) John Hume with whom we were actually in discussion (ICE-IRE:S2A- 025\$A)

However, while pied-piping is generally easier to process than stranding, depending on the type of prepositional phrase (PP) involved, in English there are also cases in which stranding can offer processing advantages. If the verb and the preposition are closely associated with each other and the preposition facilitates the interpretation of the verb, stranded prepositions appearing in close proximity to the corresponding verb can lead to a reduction in processing cost (Hawkins 1999: 260, fn. 15; Pullum & Huddleston 2002: 629; Hoffmann 2011: 59). Therefore, stranding (10a) is strongly preferred over pied-piping (10b) for prepositional verbs such as *deal with*, which can be assumed to be stored as chunks (Hoffmann 2011: 155; Jach 2021: 355). In contrast, respect, manner and degree PPs are assumed to almost categorically lead to pied-piping (13b) because they ‘do not add thematic participants to a predicate’ (Hoffmann 2011: 141). As a result, stranding (13a) would be uninterpretable. However, as Hoffmann (2011: 65–72, 155) illustrated, a strict two-way complement-adjunct distinction fails to



take into account a number of PP types with more moderate preposition placement tendencies (also see Johansson & Geisler 1998: 7; Trotta 2000: 182–4). For instance, accompaniment PPs, which also frequently involve lexically entrenched chunks such as *work with*, mildly prefer stranding (11a) over pied-piping (11b) (Hoffmann 2011: 157). In contrast, prototypical adjunct PPs, such as time PPs, whose interpretation is independent of the verb (Quirk *et al.* 1985: 511; Hoffmann 2011: 69), strongly disfavor stranding (12a) but exhibit a weaker pied-piping (12b) preference than respect, manner and degree PPs (Hoffmann 2011: 155).

- (10) (a) the sort of matter **which** I should deal **with** (ICE-IRE:S1B-066\$A)
- (b) the regime **with which** we are dealing (ICE-GB:S2B-014 #33:1:B)
- (11) (a) the people **who** I work **with** (ICE-CAN:W1B-020#83:5)
- (b) The students **with whom** she worked (ICE-CAN:W1B-027#66:2)
- (12) (a) tea **which** we could clear off **at** (ICE-GB:S1A-005 #102:1:B)
- (b) the time **at which** the data were collected (ICE-NZ:W2A-031#122:1)
- (13) (a) ? the way **which** you live your life **in**
- (b) the way **in which** you live your life (ICE-NZ:S1B-045#26:1:I)

Furthermore, pied-piping is particularly associated with the cognitively most demanding constructions (Gries 2002; Hoffmann 2011: 93–8). First of all, PPs embedded in NPs favor pied-piping more than PPs contained in verb phrases (VPs) or adjective phrases (AdjPs) because constructions with NP-embedded PPs are extremely hard to process (Trotta 2000: 184–5; Hoffmann 2011: 84–93). For instance, in (14a) the speaker cannot establish a structural relation between the filler and the gap after parsing the verb *be*. The filler can only be integrated after the human processor has encountered the NP *principal conductor*. The stranded example (14b) is even more complex as it additionally requires the parsing of the stranded preposition *of*, which itself is contained in the NP *some excerpts* (examples in (14) based on Hoffmann 2011: 86; also see Hawkins' (1999: 278) 'Principle of Valency Completeness').

- (14) (a) the orchestra, [**of which**]<sub>i</sub> he is [principal conductor<sub>i</sub>]<sub>NP</sub> (ICE-GB:W2B- 008 #71:1)
- (b) the quality of life survey [**which**]<sub>i</sub> we'll be showing you [some excerpts [**of**]<sub>PP</sub>]<sub>NP</sub> (ICE-NZ:S2B-050#28:1:C)

Another processing-related factor concerns the bridging structure, i.e. the material standing between the filler and the gap. As longer and more complex bridging structures require the parsing of additional material before the filler can be matched with the gap (Hawkins 1999: 251; 2004: 201; Trotta 2000: 188), such constructions favor preposition placement variants that are easier to process (Gries 2002: 237). Finally, Gries (2002: 237–8) also found that preposition stranding is dispreferred with the passive. Arguing that the passive is harder to process than the canonical active, he also related this effect to processing factors.

### 3 Data and methods

#### 3.1 Hypotheses

As the above has shown, the constructional competition between preposition stranding and pied-piping is strongly driven by processing factors and formality effects. In order to assess how these factors are at work in World Englishes, the following hypotheses will be tested:

- I. Processing factors should affect all varieties. Thus, all varieties should prefer prototypical constructions (such as pied-piping with adjunct PPs or NP-contained PPs; stranding with lexically entrenched V + P strings).
- II. The preference for prototypical constructions should correlate with Schneider's Dynamic Model, with Phase III exhibiting the strongest and Phase V the weakest processing effects.
- III. The strength of formality effects should correlate with the Dynamic Model, with more advanced varieties exhibiting stronger formality effects than varieties at lower stages.

#### 3.2 Data and data extraction

The data used to test these hypotheses come from the twelve ICE components for which spoken and written data are available (see [table 1](#)). Even though each ICE corpus consists of only 1 million words (except ICE-East Africa) and can thus be considered rather small, the ICE corpora include a variety of spoken and written text types and are thus ideal for an investigation of stylistic effects. Syntactic parsing of stranded prepositions being problematic, the study opted for a semi-automatic data extraction approach. After removing the extra-corpus material, the TreeTagger software (Schmid 1994) was used to tag the corpora with part-of-speech (POS) information according to the BNC Basic Tagset. Then, an R script was created that queried the tagged corpora (in the format 'word\_POS\_lemma') for all WH-words tagged as WH-determiners, WH-adverbs or WH-pronouns, using the regular expression '\\w+(DTQ|AVQ|PNQ)[^ ]+'. As the TreeTagger does not always correctly distinguish between WH-adverbs and subordinating conjunctions, additionally all occurrences of *when/where* tagged as subordinating conjunctions were extracted (using the regular expression '([Ww][Hh][Ee][Rr][Ee][Ww][Hh][Ee][Nn])\_CJS[^ ]+'). This yielded 187,158 hits, which were uploaded to the application The Red Hen Rapid Annotator (<https://beta.rapidannotator.org>; as described by Uhrig (2022)). Then, the author and one student assistant manually went through all hits to identify all relevant RCs with pied-piped or stranded prepositions. The student assistant received intensive training and a detailed coding manual. WH-words belonging to untranscribed text, editorial comments or normative insertions, and RCs in which a major part of the utterance is marked as unclear or missing (e.g. *This is certain thing that party is supposed to stand before for which that they* <O> one or two words </O> (ICE-IND:S1B-052#17:1:A))



Table 2. *Raw frequencies (percentages) of stranded and pied-piped tokens across variety stages*

	Stranding		Pied-piping		Row total	
III	263	(10.63%)	2,210	(89.37%)	2,473	(100%)
IV	154	(16.04%)	806	(83.96%)	960	(100%)
V	243	(12.06%)	1,772	(87.94%)	2,015	(100%)
Column total	660	(12.11%)	4,788	(87.89%)	5,448	(100%)

were disregarded. Moreover, RCs in which the WH-word is extracted out of subject NPs (e.g. *Montreal P Q activists some of whom threatened to resign* (ICE-CAN: S1B-021#85:2:B)) as well as clauses in which the WH-words acts as the subject of a passive clause (e.g. *an artificial barrier which must be dealt with* (ICE-CAN: W2E-008#18:1)) were excluded as they do not license variable preposition placement (Huddleston *et al.* 2002: 1093; Pullum & Huddleston 2002: 627). Finally, tokens with resumptive pronouns (*ten suggestions which I do not intend to repeat each and every one of them* (ICE-HK: S2B-022#102:2:A)), double prepositions (e.g. *the important principles for which India uh is known for* (ICE-IND: S1B-014#56:1:A)) or extra prepositions (e.g. *a new product strategy of which you'll be seeing this this afternoon* (ICE-SIN: S2A-055#59:1:A)) were discarded. This yielded a total of 5,448 stranded and pied-piped RCs (see table 2).

### 3.3 Annotation

#### 3.3.1 Linguistic factors

Following previous multifactorial studies (Gries 2002; Hoffmann 2011), all tokens were coded for the linguistic factors outlined in table 3.

The variable PREPOSITION captures idiosyncratic preferences of individual prepositions (e.g. Quirk *et al.* 1985: 664, 1253; Johansson & Geisler 1998: 75, 77). Although it would have also been interesting to investigate the idiosyncratic effects of

Table 3. *Linguistic variables investigated in the present corpus study*

Linguistic factors	Levels
PREPOSITION_PLACEMENT	stranding, pied-piping
PREPOSITION	<i>in, for, under</i> , etc.
PHRASE_TYPE	VP, AdjP, NP
PP_TYPE	lexicalized, complement_like, adjunct_like
VOICE	active, passive
COMPLEXITY	number of words between WH-word and gap
FREQUENCY_PREPOSITION	relative frequencies of prepositions computed for each individual ICE corpus
RESTRICTIVENESS	restrictive, non-restrictive

*who* and *whom* across variety types (e.g. Quirk *et al.* 1985: 370), the effect of filler types was not explored in this study. *Which* being by far the most frequent filler in RCs, the data contained only relatively few observations for other filler types (see table A1 in the Appendix), and the almost categorical preposition placement preferences of *who* (stranding) and *whom* (pied-piping) would have introduced many low frequency cells. Moreover, this factor is correlated with other variables such as PP\_TYPE.

All other variables investigated are related to processing effects (see section 2.4). With respect to PP\_TYPE, the tokens were initially annotated according to Hoffmann's (2011: 116) fine-grained classification of PP types (see table 4). Since the data contained only few observations for many levels (particularly with regard to the interaction VARIETY\_TYPE \* PP\_TYPE \* COMPLEXITY; see section 3.4) and many PP types exhibit very similar preposition placement preferences, the variable was ultimately recoded into only three levels (see table 3) to avoid high standard errors and collinearity. The 'lexicalized' level includes the PP types which are most likely to be stored as lexicalized chunks (Hoffmann 2011: 155–7). This also includes obligatory complements with *be* because Hoffmann (2011: 139) assumes 'a lexically stored constraint ... that requires *be* to co-occur with stranded prepositions only'. The second level is labeled 'complement\_like' as it contains optional complements and obligatory complements without *be*, but also movement and accompaniment PPs, which have been found to exhibit complement-like preposition placement preferences (Hoffmann 2011: 155–7). The final group 'adjunct\_like' includes all remaining adjunct PPs as well as subcategorized PPs, i.e. PPs which require a particular type of preposition but frequently have a locational meaning and exhibit adjunct-like pied-piping preferences (Hoffmann 2011: 68, 157). Although Hoffmann (2011: 160) found a categorical pied-piping preference for respect, manner, degree and frequency PPs, the present data also contained four stranded respect PP tokens (see example (15)). This indicates that stranding with these PP types is not entirely impossible in World Englishes.

- (15) Again we call into question my Lady the account given by the complainant **which the officer said in** that my Lady if one is robbed one recognises and knows the person who robbed you (ICE-JA:S2A-064#26:2:B)

Thus, it was decided to keep these PPs and to group them with the other adjunct tokens to reduce multicollinearity. Since the study investigates RCs, i.e. the most complex clause type, it is expected that all adjunct PPs strongly favor pied-piping (also see Jach 2021: 359).

COMPLEXITY was operationalized as the number of words between the filler and the stranded preposition (Szmrecsanyi 2004). For pied-piped prepositions, the earliest position in which a stranded preposition can occur was reconstructed (adapted from Hoffmann 2011: 97). The factor FREQUENCY\_PREPOSITION was included because an association between stranding and high-frequency prepositions has been suggested before but still lacks a clear explanation (Quirk *et al.* 1985: 664; Pullum & Huddleston 2002: 631; Jach 2021: 358–9). To compute the relative frequencies of prepositions for each ICE corpus, first, all non-corpus material was removed. Then, the

Table 4. *Levels of PP\_TYPE (based on Hoffmann 2011: 116)*

Factor level	PP type (token numbers)	Examples
lexicalized	VXP (73)	dinner on Friday night which I'm really looking forward to (ICE-GB:W1B-013 #89:2)
	prepositional X (682)	the policies which they stand for (ICE-NZ:S1B-024#132:1:S) the literary allegories which scholars were so fond of (ICE-IRE:S2B-032\$C)
complement_like	obligatory complement with <i>BE</i> (24)	the box which the winch is in (ICE-NZ:S1A-052#163:1:B)
	obligatory complement without <i>BE</i> (57)	the place which he lives in (ICE-GB:S1A-071 #101:1:D)
	optional complement (1112)	India whose ambassador I talked to (ICE-GB:S2B-010 #49:1:B)
	movement (168)	places where I went to (ICE-IRE:S1A-100\$B)
adjunct_like	accompaniment (65)	the people who I work with (ICE-CAN:W1B-020#83:5)
	subcategorized PP (85)	the so-called Corona, in which part of the crown or skull of St Thomas was kept (ICE-GB:W2B-003 #91:1)
	affected location (66)	the paper upon which they are writing (ICE-GB:W2D-006 #40:1)
	direction (21)	the direction from which the tram would approach (ICE-CAN:W2F-012#53:1)
	location (position) (395)	the room in which they are working (ICE-GB:W2D-006 #93:1)
	time (position in time) (268)	the moment at which Art chose life (ICE-NZ:W2F-006#130:1)
	duration/frequency (136)	the time for which each part is flooded (ICE-IRE:W2B-030\$C)
	means/instrument (364)	the tools with which they were supposed to be working (ICE-NZ:W2B-009#86:1)
	agent (2)	the person by whom he was killed
	cause/reason/ purpose/result (149)	The reasons for which we neglect taking our medicines (ICE-GB:S2B-038 #72:2:A)
	respect (1139)	the circumstances in which that initiative came to an end (ICE-NZ:S2B-041#127:1:P) a poem in which the narrator has a distinctive personality (ICE-CAN:W1A-014#32:1)
	manner (484)	the way in which they developed (ICE-GB:S1B-016 #160:1:C)
	degree (158)	the extent to which English displaced Latin (ICE-CAN:W2A-008#51:1)

Table 5. *Kappa scores for PHRASE\_TYPE, RESTRICTIVENESS and VOICE*

Variable	Cohen's kappa
PHRASE_TYPE	0.91
RESTRICTIVENESS	0.81
VOICE	0.92

Table 6. *Extralinguistic variables investigated in the present corpus study*

Extralinguistic factors	
VARIETY_TYPE	III, IV, V
TOPIC	personal_less_formal, other_less_formal, more_formal

tagged versions of the corpora (see section 3.2) were queried for all prepositions with the help of an R script using the regular expression ‘([\w\\.] +(PRP|PRF)[^ ]+)|(upto\_NN1\_upto)’.

The variables PHRASE\_TYPE, VOICE and RESTRICTIVENESS were annotated by five student assistants. Every variable was annotated by two coders. In order to ensure the comparability of the codings, training samples (N = 450 for PHRASE\_TYPE, N = 200 for VOICE and RESTRICTIVENESS) were coded and an interrater reliability analysis was conducted based on Cohen's kappa. As shown in table 5, the kappa scores for all variables are  $\geq 0.81$ . According to Landis & Koch (1977: 165), this indicates ‘almost perfect’ agreement.

3.3.2 *Extralinguistic factors*

With regard to the extralinguistic variables (see table 6), the factor VARIETY\_TYPE aims to detect effects related to the Dynamic Model. Note that the variable VARIETY could not be included in the regression analysis because of low token numbers of the individual varieties and problems with rank deficiency.

Furthermore, the factor TOPIC warrants some additional comments. Bohmann (2019: 194) reports that varieties at earlier Dynamic Model stages ‘tend towards linguistic patterns that are more formal and informational, and less affective and involved than their phase 5 ... counterparts’. As such differences in style cannot be fully captured by a strict spoken–written or formal–informal dichotomy (Bohmann 2019: 115), apart from text types, this study also takes into account topics to implement a more nuanced formality classification. Using the R package topicmodels (Grün & Hornik 2011), a topic model was fitted to assign topics to all ICE (sub)text<sup>3</sup> files.

<sup>3</sup> As many ICE text files consist of more than one subtext, initially, all ICE files were split into subtexts to avoid topic assignments that are very broad. (Exceptions are ICE-Nigeria, which includes only one text per file, and ICE-East

In order to fit a topic model, the user has to specify the number of topics to be modeled. Then, Latent Dirichlet allocation assigns each word in the data randomly to one of the topics. These topic assignments are iteratively updated, taking into account the topic assignments of all other words (Steyvers & Griffiths 2007: 436). Finally, the most likely topics for each text file as well as the most likely terms for each topic can be obtained from the fitted model (Grün & Hornik 2011: 11). This allows the development of topic labels and the assignment of a topic to each text file. For the present study, a model with twenty topics was accepted as the best result. Even though this model produced some semantically related topics, it yielded better results than models with fewer topics which did not distinguish well between individual text files.

These topics were then categorized as ‘informal’, ‘medium’ and ‘formal’ in order to explore the correlation between topics and ICE text types. Following Koch & Oesterreicher’s (2012: 450) model of ‘language of immediacy’ and ‘language of distance’, the personal topic is considered the most informal topic. Formal topics include the most informational and abstract topics such as science, while the medium level consists of topics which are informational but which also affect speakers personally to a certain extent. For instance, with regard to education, it turned out that speakers do not just present abstract information but that they also relate personal experiences in terms of schooling, for example. Thus, they were assumed to be characterized by a more involved style than the most abstract topics.<sup>4</sup> An overview of all topics categorized according to formality (‘informal’ – ‘medium’ – ‘formal’) can be found in [table A2](#) in the Appendix.

While the topics derived from the topic model and the ICE text types turned out to be largely correlated, with more formal topics such as science occurring in more formal text types, and the personal topic dominating the more informal categories, there were also some noteworthy exceptions: for instance, it became apparent that the private correspondence texts in ICE-India mainly involve administrative issues, which makes them more formal than the private correspondence files of most other ICE corpora, which tend to be dominated by personal topics (see figure A1 in the Appendix). Hence, in a first step, it was decided to use both topics and ICE text types to categorize the data into the two levels ‘more formal’ and ‘less formal’. This approach is considered superior to defining formality based on the ICE text types only because it takes into account differences that exist between the various ICE varieties with respect to topic and formality despite the shared text categories. Consequently, private correspondence files dealing with formal topics were assigned to the more formal level, whereas social letters about other topics were classified as less formal. In addition, public dialogues and unscripted monologues about personal topics were categorized as less formal, whereas files from the same categories about other topics (e.g. parliamentary debates about the economy) were considered more formal. This

Africa and ICE-Ireland because in these corpora the different design of the text unit markers does not always allow the assignation of individual extracted tokens to subtexts.)

<sup>4</sup> This classification is admittedly somewhat ad hoc and may need revision in future studies.

Table 7. *Levels of the variable TOPIC (for a detailed overview of all topics categorized according to formality see table A2 in the Appendix)*

Personal_less_formal	Other_less_formal	More_formal
Texts about personal topics from the categories: <ul style="list-style-type: none"><li>• Private correspondence</li><li>• Private dialogues</li><li>• Public dialogues</li><li>• Unscripted monologues</li></ul>	<ul style="list-style-type: none"><li>• Private dialogues about non-personal topics</li><li>• Private correspondence about medium topics</li></ul>	<ul style="list-style-type: none"><li>• Private correspondence about formal topics</li><li>• Public dialogues/unscripted monologues about non-personal topics</li><li>• Regardless of topic:<ul style="list-style-type: none"><li>- Scripted monologues</li><li>- Broadcast news</li><li>- Non-professional writing</li><li>- Business correspondence</li><li>- Academic writing</li><li>- Non-academic writing</li><li>- Reportage</li><li>- Instructional writing</li><li>- Persuasive writing</li><li>- Creative writing</li><li>- Only ICE-East Africa: Written as spoken, Legal presentations</li></ul></li></ul>

decision is based on the fact that, in terms of formality, public dialogues and unscripted monologues occupy an intermediate position between the very involved private dialogues and the very informational scripted monologues (Xiao 2009: 436–7; Bohmann 2019: 107–8). At the same time, some text types (see table 7) were categorized as more formal regardless of topics because they represent edited texts or they have been shown to be generally very informational and elaborate in style (Xiao 2009: 436–7; Bohmann 2019: 107–8). In a second step, all texts classified as less formal were then further subdivided according to topics in order to account for the fact that speakers of less advanced varieties receive less input in connection with personal topics than speakers of varieties at higher stages (Van Rooy *et al.* 2010: 346). As shown in table 7, the final variable TOPIC thus consists of the three levels ‘personal\_less\_formal’, ‘other\_less\_formal’ and ‘more\_formal’. A more fine-grained distinction between topics was not implemented to reduce multicollinearity. A simple two-way topic distinction (‘personal’ vs ‘other’) was not considered sufficient because personal topics in RCs occur almost exclusively in less formal text types. The three-level distinction thus ensures that a potential effect of the personal topic is not due to its correlation with informal text types.

3.4 Statistical analysis

In order to explore the effects of the various variables on the choice between preposition stranding and pied-piping, a generalized linear mixed-effects model (Baayen 2008:



241–302) was fitted with the help of the R package lme4 (Bates *et al.* 2015).<sup>5</sup> The statistically significant effects identified by the model will be taken as an indicator for the entrenchment of stranded and pied-piped RC constructions in the different variety types (Hoffmann 2011: 265). However, as corpora contain aggregated data from several speakers, it is important to note that they can never directly reflect cognitive entrenchment. Instead, corpora can only provide evidence for the conventionalization of constructions in a speech community (Schmid 2020: 217).

Treatment coding was applied, which means that each level of a categorical variable is compared to a specific reference level. To account for idiosyncratic preferences of individual speakers and lexical effects of prepositions, random intercepts were included for the variables FILE\_ID and PREPOSITION. A more complex random effects structure was not implemented to avoid convergence problems.<sup>6</sup> With respect to the fixed effects, all numeric variables were log-transformed, centered and standardized. Initially, a maximal model was created that included all theoretically motivated predictors, namely TOPIC, the linguistic factors described in section 3.3.1 as well as their interactions with VARIETY\_TYPE. Additionally, the three-way interaction VARIETY\_TYPE \* PP\_TYPE \* COMPLEXITY was included to account for a potential correlation between PP\_TYPE and COMPLEXITY. No other interactions were included to avoid overfitting. This model was then simplified in a backward elimination process (Zuur *et al.* 2009; Gries 2021). Likelihood ratio tests (LRT) were performed to identify which variables significantly improved the model fit. In this way, first non-significant random effects and then non-significant fixed effects were removed. The final model includes random intercepts for FILE\_ID and PREPOSITION, the predictors VOICE and PHRASE\_TYPE, as well as interactions of VARIETY\_TYPE and RESTRICTIVENESS, VARIETY\_TYPE and TOPIC, and PP\_TYPE and COMPLEXITY. None of the other predictors or interactions turned out to be significant.

The classification accuracy of the final model is 95.06 percent (precision = 95.38%, recall = 99.19%), which is highly significantly better than a baseline model that always predicts the most frequent choice, i.e. pied-piping ( $p < 0.001$ ). The index of concordance ( $C = 0.98$ ) also suggests that the model has excellent predictive capacities (as  $C > 0.8$ ; Baayen 2008: 204). The marginal  $R^2 = 0.51$  and the conditional  $R^2 = 0.74$ . Moreover, all variance inflation factors (VIFs) are well below 10 (Montgomery & Peck 1992), and the condition number  $K$  is 11.36, which suggests that there is no harmful collinearity (Baayen 2008: 182).

<sup>5</sup> Data and code are available at [https://osf.io/q6etx/?view\\_only=3271008a475c49b79fc3a9b903739568](https://osf.io/q6etx/?view_only=3271008a475c49b79fc3a9b903739568).

<sup>6</sup> Prepositions with  $\leq 5$  observations were combined into one group ‘other’. With regard to FILE\_ID, no levels were pooled. Even though the relatively large number of FILE\_ID levels with only one observation ( $N = 1448$ ) may lead to inaccurate estimates of the random effects variance of FILE\_ID, this should not have a negative effect on the estimation of the fixed effects (Meteyard & Davies 2020).

Table 8. *Variance estimates and standard deviations of random effects*

Groups	Variance	SD
FILE_ID	1.878	1.37
PREPOSITION	1.01	1.005

4 Results

4.1 Random effects

The random intercept estimates for FILE\_ID and PREPOSITION (see table 8) show that both individual speakers and individual prepositions vary in their baseline preference for preposition stranding and pied-piping. The intercept adjustments for the various prepositions, with positive estimates indicating a stranding preference and negative estimates indicating a pied-piping preference, can be seen in figure 1. The prepositions that favor stranding most are *about*, *through* and *at*, whereas *among*, *around* and *in* have the strongest pied-piping preference.

4.2 Fixed effects

Table 9 lists the effects of the individual variables on preposition placement in RCs. The model predicts log-odds for preposition stranding. A positive *b* value indicates that a level

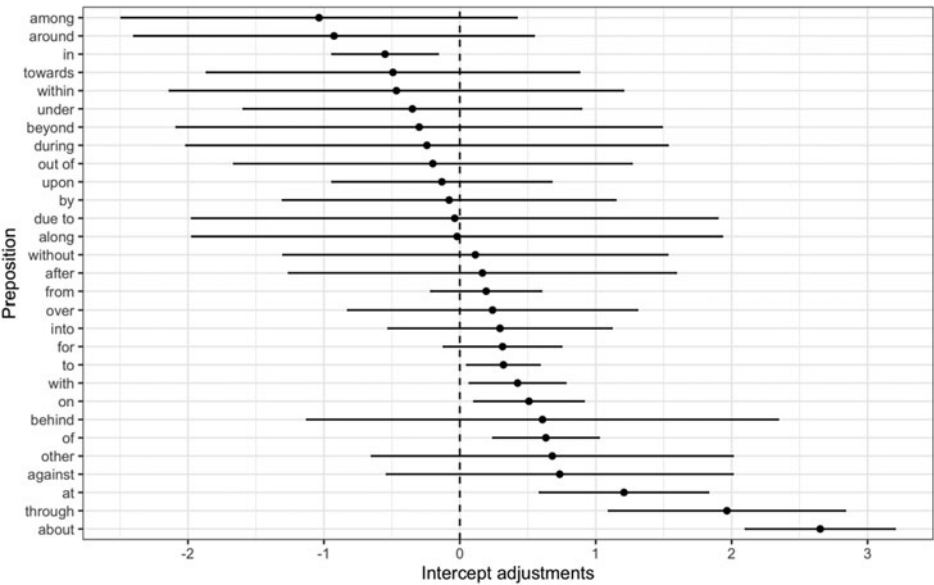


Figure 1. Intercept adjustments of PREPOSITION and Wald confidence intervals

Table 9. *Coefficients of significant fixed effects (predicted level: stranding)*

Predictors	<i>b</i>	SE	<i>z</i>	<i>p</i>
(Intercept)	-1.0192	0.3056	-3.335	<b>&lt;0.001</b>
VARIETY_TYPE				
V > III	0.2931	0.1914	1.531	0.1258
V > IV	0.6775	0.2319	2.9216	<b>&lt;0.01</b>
PP_TYPE				
lexicalized > complement_like	-0.9011	0.1956	-4.606	<b>&lt;0.001</b>
lexicalized > adjunct_like	-4.3859	0.3069	-14.2903	<b>&lt;0.001</b>
COMPLEXITY (logged)				
one-unit increase	-0.128	0.1539	-0.8313	0.4058
VOICE				
active > passive	-2.1496	0.2564	-8.3848	<b>&lt;0.001</b>
PHRASE_TYPE				
VP > AdjP	-0.3963	0.254	-1.5602	0.1187
VP > NP	-1.5498	0.3064	-5.0588	<b>&lt;0.001</b>
RESTRICTIVENESS				
restrictive > non-restrictive	0.7232	0.2629	2.7511	<b>&lt;0.01</b>
TOPIC				
more_formal > other_less_formal	2.2438	0.7807	2.874	<b>&lt;0.01</b>
more_formal > personal_less_formal	3.3642	0.418	8.049	<b>&lt;0.001</b>
PP_TYPE * COMPLEXITY (logged)				
complement_like + one-unit increase	-0.0001	0.1981	-0.0005	0.9996
adjunct_like + one-unit increase	-0.5323	0.2669	-1.9942	<b>&lt;0.05</b>
VARIETY_TYPE * RESTRICTIVENESS				
III + non-restrictive	-0.9285	0.3651	-2.543	<b>&lt;0.05</b>
IV + non-restrictive	-0.9042	0.4786	-1.8895	0.0588
VARIETY_TYPE * TOPIC				
III + other_less_formal	-0.7645	0.9109	-0.8393	0.4013
IV + other_less_formal	-0.1168	0.9789	-0.1193	0.905
III + personal_less_formal	-1.8457	0.6594	-2.7988	<b>&lt;0.01</b>
IV + personal_less_formal	-0.3074	0.7376	-0.4168	0.6768

change of a categorical predictor or a one-unit increase of a numeric variable lead to an increasing probability of preposition stranding, while a negative *b* value indicates that preposition stranding becomes less likely.

According to the model, preposition stranding is negatively associated with the passive. Likewise, PPs contained in NPs lead to a decreasing probability of preposition stranding, whereas there is no significant difference between PPs embedded in VPs and AdjPs.

With regard to the other predictors, it is important to keep in mind that they are involved in interactions. While complement\_like and particularly lexicalized PPs favor pied-piping more than adjunct\_like PPs, not all PP types are affected equally by increasing complexity. As shown in table 9, a one-unit increase in complexity (logged) makes preposition stranding significantly less likely for adjunct\_like PPs

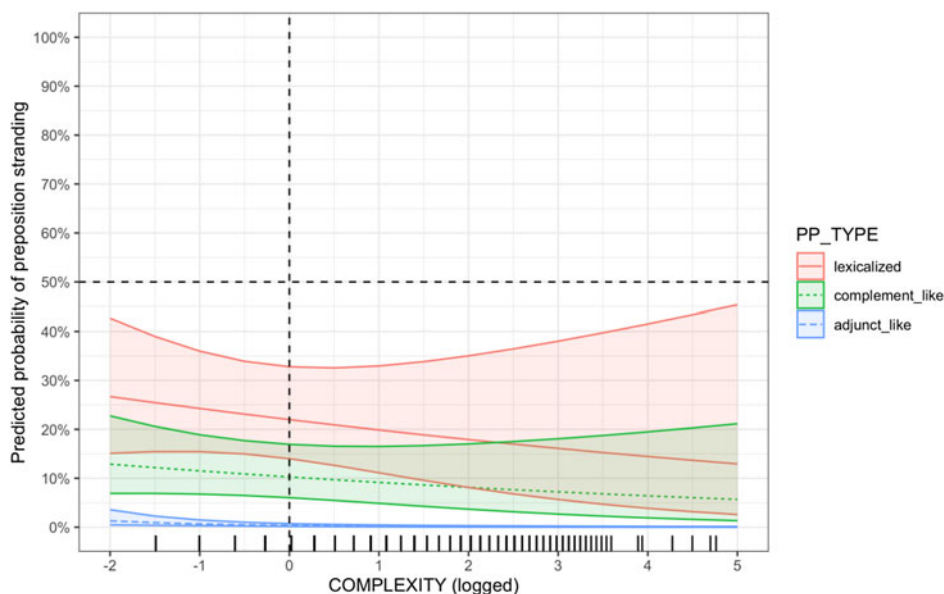


Figure 2. The effect of COMPLEXITY (logged) \* PP\_TYPE

than for lexicalized PPs. This effect is also represented in figure 2: pied-piping is generally preferred in all RCs regardless of complexity or PP type. However, compared to adjunct\_like PPs, lexicalized and complement\_like PPs have a higher probability of preposition stranding when the clause is less complex, while increasing complexity leads to a stronger pied-piping preference. In contrast, adjunct\_like PPs are not even stranded in the simplest RCs.

Turning to VARIETY\_TYPE, this predictor is also involved in two interactions. First, not all variety types are affected equally by RESTRICTIVENESS. Only in the Stage-V varieties does a change from restrictive to non-restrictive relative clauses lead to a significant increase in stranding. In contrast, RESTRICTIVENESS hardly influences the constructional choices of speakers of Stage-III and Stage-IV varieties. Instead, as shown in the corresponding effects plot (see figure 3), in both restrictive and non-restrictive RCs, Phase III exhibits a pied-piping probability that is comparable to that of restrictive RCs in Phase V, while in Phase IV both restrictive and non-restrictive RCs show a pied-piping preference that is at a similar level as that of non-restrictive RCs in Phase V.

Second, TOPIC has a much stronger effect on variety types IV and V. The significant negative coefficient of the interaction in table 9 shows that, compared to Stage V, speakers of Stage-III varieties are significantly less likely to use a stranded preposition in the personal\_less\_formal level than in the more\_formal level. In contrast, Phase IV patterns with Phase V. As illustrated in figure 4, both Stages IV and V display an almost categorical pied-piping preference in more formal contexts, but exhibit an

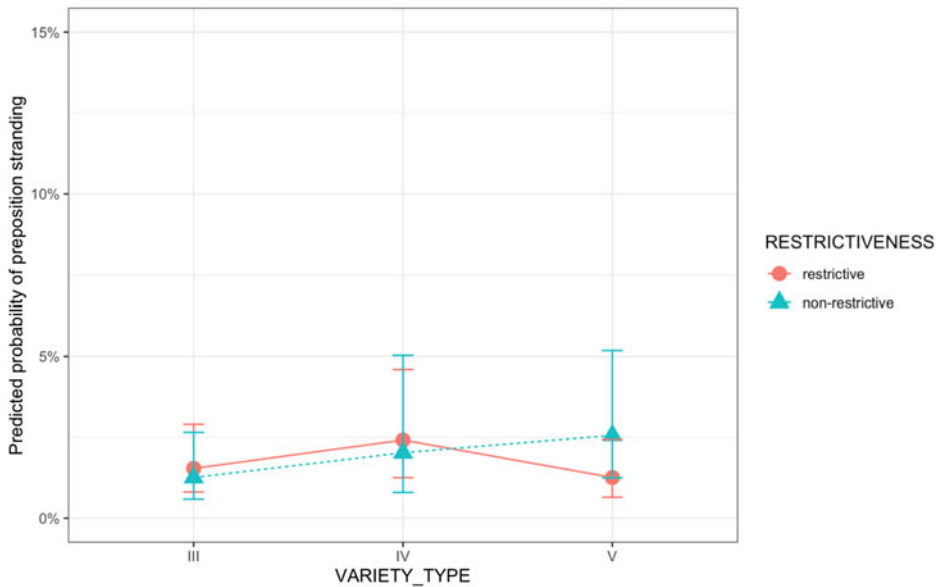


Figure 3. The effect of RESTRICTIVENESS \* VARIETY\_TYPE

increasing probability of stranding in less formal contexts dealing with non-personal topics and the strongest stranding preference in less formal texts about personal topics. At the same time, Phase III only shows a slightly weaker pied-piping preference in less

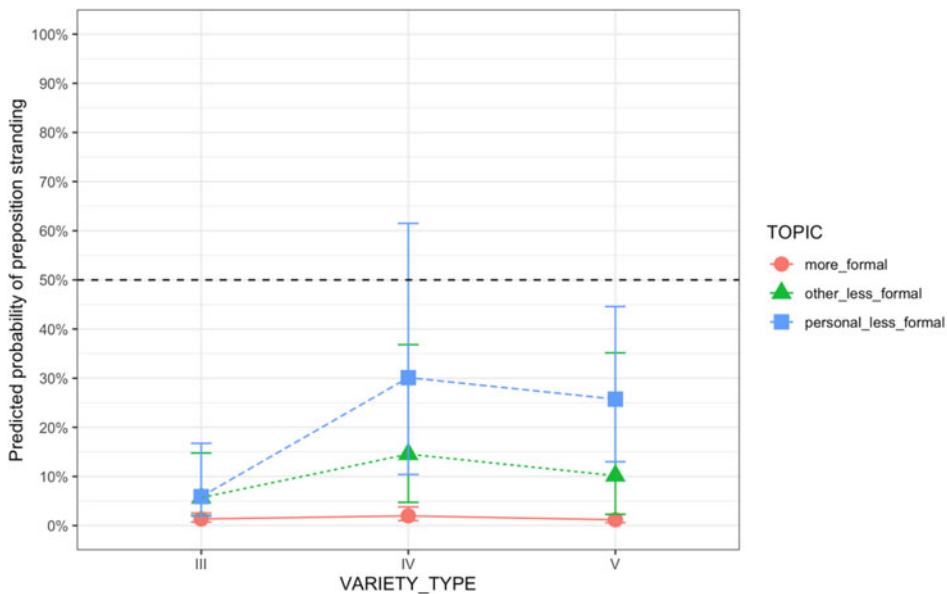


Figure 4. The effect of TOPIC \* VARIETY\_TYPE

formal text types than in more formal ones, and the distinction between personal and non-personal topics does not make a difference at all.

## 5 Discussion

The study was based on the assumption that processing factors should affect all variety types (Hypothesis I), but that the strength of processing effects (Hypothesis II) and formality effects (Hypothesis III) should correlate with Dynamic Model stages.

Hypothesis I was confirmed because the model shows that the constructional choices of all variety types are strongly driven by processing factors. Speakers disprefer variants that involve an increase in processing efforts such as stranded prepositions contained in NP-embedded PPs, which cannot be parsed upon encountering the main verb of the clause. Moreover, this study confirms Gries' (2002) finding that the probability of preposition stranding increases in active constructions. That passive constructions are associated with higher processing demands (Rohdenburg 1996: 174) receives support from psycholinguistic studies (Davison & Lutz 1985) and could be due to the fact that the semantic roles in passive constructions are expressed in a different syntactic order than in the more frequent active construction (Diessel 2004: 14; Wanner 2009: 14). Finally, the effect of PP TYPE can also be linked to processing factors. Speakers favor pied-piping most strongly with adjunct PPs because separating prepositions with predicate-independent meanings from their complements is associated with increased processing cost. At the same time, PP types which are more likely to be stored as V + P chunks are stranded more frequently, even though pied-piping still remains the preferred option. That even these PP types generally favor pied-piping over stranding (see figure 2) could be the result of formality because speakers might consciously avoid stranding with, for example, prepositional verbs in more monitored and more formal situations. The three-way interaction PP\_TYPE \* TOPIC \* VARIETY\_TYPE was, however, too complex to be tested in this study.

With regard to complexity, the logistic regression analysis also points to an effect that is not in line with previous studies. Contrary to the predictions of Trotta (2000: 188) and Hoffmann (2011: 168, 224–5), who claimed that stranding should be strongly preferred with V + P chunks in longer FGDs, this study showed that the stranding probability for such PP types decreases with increasing complexity. Although Hoffmann and Trotta correctly point out that in very long constructions breaking up lexicalized verb-preposition strings could lead to an increase in processing efforts, the great complexity resulting from long FGDs (cf. Hawkins' (1999: 251; 2004: 27) 'Minimize FGDs' principle) and PP-contained gaps might lead to an avoidance of pied-piping in such complex environments. Complex RCs with, for example, prepositional verbs being an infrequent phenomenon (also see the large confidence band in figure 2), one reason for the difference between the results of this study and Hoffmann's (2011) analysis could be that Hoffmann's data, which came from only two ICE-corpora, may not have contained as long RCs as those used in this study. Furthermore, compared to



Hoffmann's data, the complex RCs included in the present study may appear in more formal registers and may thus be more strongly associated with pied-piping. A thorough investigation of the interaction of complexity and PP type thus requires a larger database or complementing experimental studies.

With regard to hypotheses II and III, no direct correlation between Dynamic Model stages and the strength of processing and stylistic effects could be confirmed because Phase IV turned out to pattern with Phase V. In some contexts, Stage IV even favors the non-prototypical stranding RC construction more than the most advanced varieties. However, while processing factors seem to influence the constructional choices of speakers of all variety stages in a very similar way, there is at least some evidence for the fact that they have the strongest effect on varieties at lower stages of Schneider's model. In the least complex clause type investigated, i.e. non-restrictive RCs, Stage III displays the strongest preference for the prototypical pied-piped construction. At the same time, surprisingly, Phase IV uses even more stranded prepositions than Phase V in restrictive RCs, i.e. the most complex environment. Since this cannot be explained from a processing perspective, this stronger stranding preference of Jamaican English (JamE) and Singapore English (SgE) could be due to L1 influence. While a detailed analysis of potential sources of L1 transfer is beyond the scope of this article, a brief review of the L1s spoken in Singapore indicates that particularly speakers of Chinese languages and Malay, which make up a substantial proportion of the L1s spoken in Singapore (Department of Statistics Singapore 2021: 23), do not encounter pied-piped RC constructions in their L1 input. As shown in (12), in Mandarin Chinese elements with preposition-like functions are normally lost under relativization (example (16) from Li & Thompson 1989: 583).

- (16) páshǒu      tōu      dōngxī      de      fāngfǎ  
 pickpocket steal thing REL method  
 'the method by which pickpockets steal things'

In Malay, only subjects and direct objects can be relativized (Keenan & Comrie 1977: 71). At the same time, SgE favors *that*-relatives, in which stranding is obligatory, over *which*-relatives (Suárez-Gómez 2015: 257–8). The combination of both factors, i.e. the absence of pied-piped RC constructions in speakers' L1s and the exposure to a high proportion of *that*-RCs, which might serve as a model for WH-relatives, could thus lead to an overall stronger stranding preference in SgE. Speakers of JamE may prefer stranding because they regularly encounter stranded RC constructions in Jamaican Creole, which licenses preposition stranding (17a) but prohibits pied-piping (17b) in RCs (examples in (17) adapted from Patrick 2004: 426).

- (17) (a) di      dort      we      shi      waak      pan  
           the ground REL she walk on  
       (b) \* di      dort      pan      we      shi      waak  
           the ground on REL she walk  
       'the ground which she walks on'

Now, of course, it is important to keep in mind that there are also Stage-III varieties in whose L1s preposition pied-piping in RCs is not entrenched. For instance, the Bantu languages spoken in Kenya and Tanzania allow neither stranding nor pied-piping but require a resumptive pronoun if a prepositional object is relativized (Riedel 2010: 218). However, in contrast to the Stage-IV varieties, in at least some Stage-III varieties, substrate effects could favor pied-piping. Examples of L1s in which pied-piping is the norm in RCs are the Indo-Aryan languages spoken in India, such as Hindi (Keenan & Comrie 1979: 338). However, as the logistic regression analysis did not allow the investigation of variety-specific preferences, no definite conclusions about L1 effects can be drawn from this study.

In sum, disregarding the higher preference for stranding in non-restrictive RCs in Phase III, the study confirms previous research (see section 1) that found that processing factors drive the constructional choices of speakers in very similar ways in varieties of English around the world. But why is it that KenE speakers in Hoffmann's (2011) study rated stranding with prototypical adjunct PPs lower than BrE speakers, whereas this study did not identify a significant interaction between PP type and variety type? One reason behind this may be that Hoffmann's results were based on magnitude estimation experiments, while this analysis relied on corpus data. The corpus data used in this study simply may not have contained more stranded adjunct PP tokens produced by more advanced speakers. At the same time, the experimental data may have produced different results because '[i]t is not news to say that people will say one thing and do another' (Labov 1975: 104). This underlines the importance of regarding introspective and corpus data as 'corroborating evidence' (Hoffmann 2006: 167).

Three other reasons are suggested why not more processing-related variables turned out to have a stronger effect on varieties at lower stages of Schneider's evolutionary cycle. First, this study focused on preposition placement in RCs only, i.e. the most complex clause type, in which not only a relation between the filler and the gap but also between the antecedent noun and the relativizer has to be established. In such highly complex contexts even the most advanced speakers might prefer variants that involve a reduction in processing efforts more than in, for example, interrogative clauses, in which only the filler and the gap have to be matched. Such an explanation is in line with Szmrecsanyi *et al.* (2016: 122), who concluded, based on a study of particle placement, that processing-related differences between varieties only arise in 'contexts where the processing load is relatively minimal'. Second, pied-piping in RCs being the prototypical choice, it can be assumed that speakers of all variety types are exposed to a lot of positive input for this construction. As a result, P + WH chunks such as *in which* or *to whom* can be assumed to become deeply entrenched in the mental grammars of all speakers (although Phase-V speakers should possess the most of these constructions since they receive most input). These entrenched partly schematic constructions can be activated easily and might contribute to prototypical constructional choices in all varieties regardless of their developmental stage. A third reason may have to do with the fact that this study investigated the correlation between the strength of processing factors and variety phases. Such a phase-related approach not only makes it difficult to identify potential L1

effects, but it is also important to keep in mind that the assignment of varieties to Dynamic Model phases as outlined in Schneider (2007: 113–250) is not based on empirical results and should thus be taken with a grain of salt. Hence, the results may also be skewed by the fact that the variable VARIETY\_TYPE ignores input differences that may exist between different varieties subsumed under the same Dynamic Model phase.<sup>7</sup>

While processing constraints proved to be highly relevant in all varieties, clear differences between variety stages were observed with respect to formality and topic, which strongly influence the constructional choices of Phases IV and V but not of Phase III. The study thus corroborates the results of Hoffmann (2011), who found a strong formality effect for BrE RCs but only a weak one for KenE. Furthermore, the results showed that speakers of the more advanced varieties do not just vary between more formal and less formal contexts but also between personal topics and more abstract topics such as language or education. Stage III, however, does not exhibit a topic effect at all. Hoffmann (2011: 267–8) attributed the absence of a formality effect in KenE to the fact that, in contrast to BrE speakers, KenE speakers do not have an entrenched stranded RC construction because they lack sufficient input for stranded prepositions. An input-based explanation can also account for the lack of stylistic variation observed in the Phase-III varieties in this study. Speakers of lower Dynamic Model stages encounter English mostly in formal situations, in which pied-piped RC constructions are much more likely to surface than stranded prepositions. In contrast, other languages are preferred in connection with the most personal matters (Van Rooy *et al.* 2010: 346). As the entrenchment of abstract constructions depends on input frequency (see section 2.2), from a usage-based perspective, it can thus be assumed that Phase-III speakers do not possess an abstract constructional template for stranded RCs, but that they only use stranded prepositions in RCs as part of specific lexicalized V + P chunks that are entrenched and can be activated easily (also see Jach 2021: 366). Also note that the fact that speakers are likely to encounter stranded prepositions in *that*- and zero-relatives does not automatically lead to the entrenchment of an abstract stranded WH-RC construction. Since *that*- and zero-relatives mostly involve lexicalized verb–preposition strings (Hoffmann 2011: 123, 128), these constructions are likely to contribute to the entrenchment of specific V + P chunks, but it is cognitively very implausible to assume that speakers store one general abstract stranded relative clause construction that accounts for stranding in WH-, *that*- and zero-relatives (for details see Hoffmann 2011: 264–6). Consequently, Phase-III speakers only possess an abstract pied-piped RC construction, which is used regardless of formality and topic. Support for this claim comes from L1 acquisition, where statistical preemption can only begin to play a relevant role after children have acquired constructional alternatives (Tomasello 2003: 180).

That Phase IV patterns with Phase V is slightly surprising but can also be linked to input. The Stage-IV varieties were also frequently exposed to the pied-piped construction in the formal domains of use through which the English language was

<sup>7</sup> I thank an anonymous reviewer and Bernd Kortmann for this suggestion.

introduced in the British colonies. This led to the entrenchment of an association between the pied-piped RC construction and formal contexts. However, since Stage IV tends to use even slightly more stranded RCs than Stage V, it is plausible to conclude that speakers of Stage-IV varieties also have an entrenched abstract stranded RC construction. As a result of statistical preemption (Tomasello 2003: 300; Goldberg 2006: 94–8), this stranded construction became associated with more informal situations (also see Hoffmann 2011: 268–9 for a similar explanation regarding BrE). Drawing on terminology from Schmid's (2020) Entrenchment-and-Conventionalization Model, this means that pragmatic associations that link the stranded RC construction to informal contexts and usage situations in which speakers recount personal experiences became routinized and entrenched in the minds of individual speakers (Schmid 2020: 208–9). As a result, the stranded construction is much more likely to be activated in informal situations than the pied-piped alternative. At the same time, the stranded construction became conventionalized as the appropriate choice in informal contexts at the community-level through repeated usage events in which these patterns of associations were activated (Schmid 2020: 5–6). The 'dynamic interaction between usage, conventionalization, and entrenchment' (Schmid 2020: 3) can thus explain how in the more advanced varieties specific preposition placement constructions became associated with specific contexts.

## 6 Conclusion

The present study used generalized linear mixed-effects modeling to explore the use of stranded and pied-piped RC constructions in twelve varieties of English that represent Phases III, IV and V of Schneider's Dynamic Model. In line with previous studies, it found that processing factors strongly affect the constructional choices of speakers of all variety stages. At the same time, contrary to the initial hypotheses, the strength of processing constraints and stylistic effects did not turn out to correlate with Dynamic Model stages. While processing factors have similar effects on all variety phases, with regard to stylistic variation, the study suggested a clear two-way distinction between Phase III and Phases IV/V, and provided a usage-based explanation for the lack of a strong formality effect in the Stage-III varieties. Furthermore, the article showed how the investigation of topics can add further nuance to the study of formality.

One reason for the fact that not more processing-related variables showed an effect of variety stage could be that this study focused on preposition placement in RCs, i.e. the most complex clause type. Studies to come should thus analyze preposition placement in other, less complex clause types in order to assess the interplay between a variety's developmental stage, processing constraints and formality effects in cognitively less-demanding constructions. Moreover, one of the limitations of this study was that the dataset did not allow a detailed examination of variety-specific preferences and potential L1 transfer effects. Preposition placement in RCs should thus also be analyzed based on data from larger corpora. Finally, experimental studies should be conducted in order to validate the conclusions drawn from this corpus study.

*Author's address:*

*Department of English Language and Linguistics*  
*Catholic University of Eichstätt-Ingolstadt*  
*Universitätsallee 1*  
*85072 Eichstätt*  
*Germany*  
[victoria.mussemann@ku.de](mailto:victoria.mussemann@ku.de)

## References

- Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.
- Bernaish, Tobias, Stefan Th. Gries & Joybrato Mukherjee. 2014. The dative alternation in South Asian English(es): Modelling predictors and predicting prototypes. *English World-Wide* 35(1), 7–31.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *The Longman grammar of spoken and written English*. London: Longman.
- Bohmann, Axel. 2019. *Variation in English worldwide: Registers and global varieties*. Cambridge: Cambridge University Press.
- Buregeya, Alfred. 2019. *Kenyan English*. Berlin: De Gruyter.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, William & D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Davison, Alice & Richard Lutz. 1985. Measuring syntactic complexity relative to discourse context. In Arnold M. Zwicky, David R. Dowty & Lauri Karttunen (eds.), *Natural language parsing: Psychological, computational, and theoretical Perspectives*, 26–66. Cambridge: Cambridge University Press.
- Dayag, Danilo T. 2016. Preposition stranding and pied-piping in Philippine English: A corpus-based study. In Gerhard Leitner, Azirah Hashim & Hans-Georg Wolf (eds.), *Communicating with Asia: The future of English as a global language*, 102–19. Cambridge: Cambridge University Press.
- Department of Statistics Singapore. 2021. *Census of population 2020: Statistical release 1. Demographic characteristics, education, language and religion*. [www.singstat.gov.sg/publications/reference/cop2020/cop2020-sr1/census20\\_stat\\_release1](http://www.singstat.gov.sg/publications/reference/cop2020/cop2020-sr1/census20_stat_release1) (accessed 10 June 2023).
- Diessel, Holger. 2004. *The acquisition of complex sentences*. Cambridge: Cambridge University Press.
- Diessel, Holger. 2019. *The grammar network: How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.
- Dubois, Tanguy, Magali Paquot & Benedikt Szmrecsanyi. 2023. Alternation phenomena and language proficiency: The genitive alternation in the spoken language of EFL learners. *Corpus Linguistics and Linguistic Theory* 19(3), 427–50.
- Gilquin, Gaëtanelle & Magali Paquot. 2008. Too chatty: Learner academic writing and register variation. *English Text Construction* 1(1), 41–61.

- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Greenbaum, Sidney & Gerald Nelson. 1996. The International Corpus of English (ICE) Project. *World Englishes* 15(1), 3–15.
- Gries, Stefan Th. 2002. Preposition stranding in English: Predicting speakers' behaviour. In Vida Samiian (ed.), *Proceedings of the Western Conference on Linguistics*, vol. 12, 230–41. Fresno, CA: California State University.
- Gries, Stefan Th. 2021. (Generalized linear) mixed-effects modeling: A learner corpus example. *Language Learning* 71(3), 757–98.
- Grün, Bettina & Kurt Hornik. 2011. topicmodels: An R package for fitting topic models. *Journal of Statistical Software* 40(13), 1–30.
- Hawkins, John A. 1999. Processing complexity and filler-gap dependencies across grammars. *Language* 75(2), 244–85.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Heller, Benedikt, Benedikt Szmeccsanyi & Jason Graffmiller. 2017. Stability and fluidity in syntactic variation world-wide: The genitive alternation across varieties of English. *Journal of English Linguistics* 45(1), 3–27.
- Hoffmann, Thomas. 2006. Corpora and introspection as corroborating evidence: The case of preposition placement in English relative clauses. *Corpus Linguistics and Linguistic Theory* 2(2), 165–95.
- Hoffmann, Thomas. 2007. 'I need data which I can rely on': Corroborating empirical evidence on preposition placement in English relative clauses. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base*, 161–84. Berlin: De Gruyter.
- Hoffmann, Thomas. 2011. *Preposition placement in English: A usage-based approach*. Cambridge: Cambridge University Press.
- Hoffmann, Thomas. 2013. Obtaining introspective acceptability judgements. In Julia Schlüter & Manfred Krug (eds.), *Research methods in language variation and change*, 99–118. Cambridge: Cambridge University Press.
- Hoffmann, Thomas. 2022. *Construction Grammar: The structure of English*. Cambridge: Cambridge University Press.
- Hoffmann, Thomas & Graeme Trousdale. 2013. Construction Grammar: Introduction. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar*, 1–12. Oxford: Oxford University Press.
- Hornstein, Norbert & Amy Weinberg. 1981. Case theory and preposition stranding. *Linguistic Inquiry* 12(1), 55–91.
- Huddleston, Rodney D. & Geoffrey K. Pullum *et al.* 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Huddleston, Rodney D., Geoffrey K. Pullum & Peter Peterson. 2002. Relative constructions and unbounded dependencies. In Huddleston & Pullum *et al.*, 1031–96.
- Jach, Daniel. 2018. A usage-based approach to preposition placement in English as a second language. *Language Learning* 68(1), 271–304.
- Jach, Daniel. 2021. Preposition placement in multilingual constructions: Something I was dealing with. In Hans C. Boas & Steffen Höder (eds.), *Constructions in contact 2: Language change, multilingual practices, and additional language acquisition*, 339–74. Amsterdam: John Benjamins.
- Johansson, Christine & Christer Geisler. 1998. Pied piping in spoken English. In Antoinette Renouf (ed.), *Explorations in Corpus Linguistics*, 67–82. Amsterdam: Rodopi.
- Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1), 63–99.



- Keenan, Edward L. & Bernard Comrie. 1979. Data on the noun phrase accessibility hierarchy. *Language* 55(2), 333–51.
- Koch, Peter & Wulf Oesterreicher. 2012. Language of immediacy – language of distance: Orality and literacy from the perspective of language theory and linguistic history. In Claudia Lange, Beatrix Weber & Göran Wolf (eds.), *Communicative spaces. Variation, contact, and change. Papers in honour of Ursula Schaefer*, 441–73. Frankfurt am Main: Peter Lang.
- Labov, William. 1975. Empirical foundations of linguistic theory. In Robert Austerlitz (ed.), *The scope of American linguistics: Papers of the first golden anniversary symposium of the Linguistic Society of America, held at the University of Massachusetts, Amherst on July 24 and 25, 1974*, 77–134. Lisse: Peter de Ridder.
- Landis, J. Richard & Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1), 159–74.
- Langacker, Ronald W. 2008. *Cognitive Grammar: A basic introduction*. Oxford: Oxford University Press.
- Li, Charles N. & Sandra Thompson. 1989. *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.
- Meteyard, Lotte & Robert A.I. Davies. 2020. Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language* 112, 104092.
- Montgomery, Douglas C. & Elizabeth A. Peck. 1992. *Introduction to linear regression analysis*. New York: Wiley.
- Patrick, Peter L. 2004. Jamaican Creole: Morphology and syntax. In Bernd Kortmann & Edgar W. Schneider (eds.), *A handbook of varieties of English: A multimedia reference tool*, vol. 2, 407–38. Berlin: De Gruyter.
- Pollard, Carl & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.
- Pullum, Geoffrey K. & Rodney D. Huddleston. 2002. Prepositions and prepositional phrases. In Huddleston & Pullum *et al.*, 597–661.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Riedel, Kristina. 2010. Relative clauses in Haya. *ZAS Papers in Linguistics* 53, 211–26.
- Rohdenburg, Günter. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7(2), 149–82.
- Van Rooy, Bertus, Lize Terblanche, Christoph Haase & Joseph Schmied. 2010. Register differentiation in East African English: A multidimensional study. *English World-Wide* 31(3), 311–49.
- Ross, John R. 1986. *Infinite syntax!* Norwood, NJ: Ablex.
- Röthlisberger, Melanie, Jason Grafmiller & Benedikt Szendrői. 2017. Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics* 28(4), 673–710.
- Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*. Manchester. [www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/](http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/) (accessed 26 April 2021).
- Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.
- Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.
- Steyvers, Mark & Tom Griffiths. 2007. Probabilistic topic models. In Thomas K. Landauer, Danielle S. McNamara, Simon Dennis & Walter Kintsch (eds.), *Handbook of latent semantic analysis*, 427–48. New York: Psychology Press.
- Suárez-Gómez, Cristina. 2014. Relative clauses in Southeast Asian Englishes. *Journal of English Linguistics* 42(3), 245–68.

- Suárez-Gómez, Cristina. 2015. Adverbial relative clauses in World Englishes. *World Englishes* 34(4), 620–35.
- Szmrecsanyi, Benedikt. 2004. On operationalizing syntactic complexity. In Gérard Purnelle, Cédric Fairon & Anne Dister (eds.), *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis*, vol. 2, 1032–9. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37(2), 109–37.
- Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Trotta, Joe. 2000. *Wh-clauses in English: Aspects of theory and description*. Amsterdam: Rodopi.
- Uhrig, Peter. 2022. Large-scale multimodal corpus linguistics – The big data turn. Habilitation thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg.
- Wanner, Anja. 2009. *Deconstructing the English passive*. Berlin: De Gruyter.
- Wulff, Stefanie, Nicholas Lester & Maria T. Martinez-Garcia. 2014. *That*-variation in German and Spanish L2 English. *Language and Cognition* 6(2), 271–99.
- Wulff, Stefanie, Stefan Th. Gries & Nicholas Lester. 2018. Optional *that* in complementation by German and Spanish learners. In Andrea Tyler, Lihong Huang & Hana Jan (eds.), *What is applied cognitive linguistics? Answers from current SLA research*, 99–120. Berlin: De Gruyter.
- Wulff, Stefanie & Stefan Th. Gries. 2019. Particle placement in learner language. *Language Learning* 69(4), 873–910.
- Xiao, Richard. 2009. Multidimensional analysis and the study of World Englishes. *World Englishes* 28(4), 421–50.
- Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. Mixed effects modelling for nested data. In Alain F. Zuur, Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev & Graham M. Smith (eds.), *Mixed effects models and extensions in ecology with R*, 101–42. New York: Springer.

## Appendix

Table A1. Raw frequencies of filler types across variety stages

Filler type	Variety stage	Raw frequency
what	III	3
	IV	1
	V	1
whence	V	3
where	III	70
	IV	16
	V	15
which	III	2,221
	IV	833
	V	1,840
who	III	22
	IV	28
	V	29
whom	III	133
	IV	75
	V	110
whose	III	24
	IV	7
	V	17

Table A2. Topics categorized according to formality

Informal	Medium	Formal
Personal	Asia	Administration
	Education	Art and literature
	Household and gardening	Court
	Language	Disease
	Situational descriptions	Economy
	Society	Government
	Sports	History and culture
		National development
		News
		Philosophy
		Science
		Technology

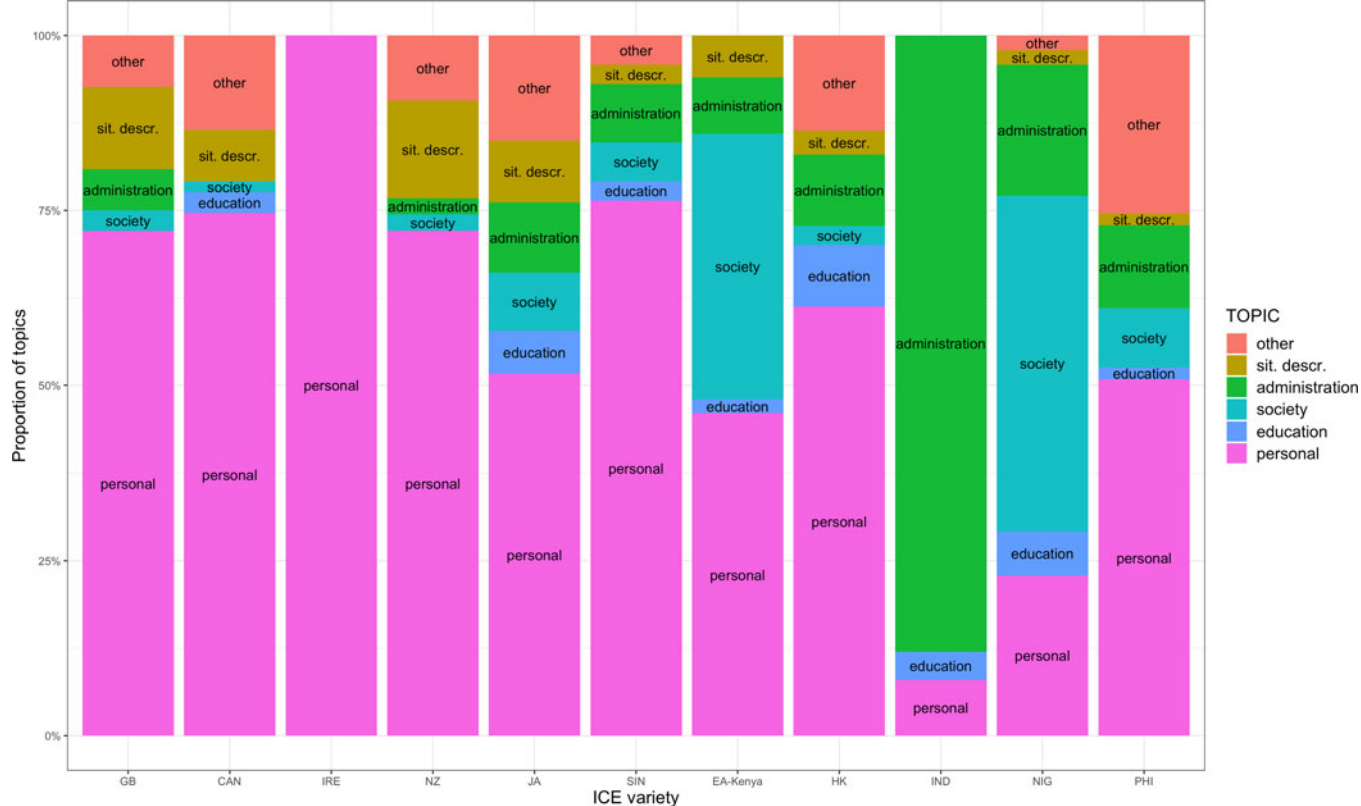


Figure A1. Topics per ICE variety in private correspondence  
Note: ICE-Tanzania is excluded because it does not contain social letters.