

Constructional Competition –
Preposition Stranding and Pied-piping in World Englishes

Inaugural-Dissertation zur Erlangung des Doktorgrades (Dr. phil.)
der Sprach- und Literaturwissenschaftlichen
Fakultät der Katholischen Universität Eichstätt-Ingolstadt

vorgelegt von

Victoria Johanna Mußemann, Eichstätt

2024

Referent: Prof. Dr. Thomas Hoffmann

Korreferent: Prof. Dr. Alexander Bergs

Tag der mündlichen Prüfung: 18.07.2024

Abstract

Adopting a combination of Usage-based Construction Grammar and New Englishes approaches, the present study explores the constructional competition between preposition stranding (*What is he talking about?*) and pied-piping (*About what is he talking?*) in 12 varieties of English at different stages of Schneider's (2007) Dynamic Model in order to gain more insights into the cognitive principles underlying the evolution of postcolonial Englishes. Analyzing 10,079 tokens from the International Corpus of English (ICE) and using mixed-effects logistic regression analysis, the study investigates whether the strength of processing factors and stylistic variables is correlated with Dynamic Model stages. Moreover, drawing on Hierarchical Configural Frequency Analysis (HCFA) and Large Number of Rare Events (LNRE) models, it tests Hoffmann's (2021) Dynamic Model Productivity hypothesis, which predicts that more advanced varieties should exhibit a higher productivity of constructional slots, while varieties at lower stages of evolution should rely more on a small set of prototypical fillers.

The results are largely in line with Hoffmann's (2011) analysis of preposition placement, which found that Kenyan English is more strongly affected by processing constraints than British English, while stylistic variables have a weaker effect in the L2 variety. Even though processing constraints turn out to strongly affect the constructional choices of speakers of all varieties, less advanced varieties are more likely to use pied-piped WH-interrogative constructions with adjunct PPs, which suggests that varieties at lower stages of evolution have a stronger preference for constructions that facilitate parsing in contexts in which the processing load is not maximal. Furthermore, in line with the frequent claim that New Englishes are characterized by a lack of register distinctions, the strength of stylistic variables turns out to be correlated with Dynamic Model stages. While Phases IV and V have formal pied-piped and informal stranded finite WH-relative clause constructions, Phase III only has the pied-piped variant. This effect can be attributed to two factors: the formal input that is typical of postcolonial Englishes, and the processing complexity of stranding constructions. That preposition stranding is less deeply entrenched in varieties at lower stages is also indicated by the fact that the slots of stranding constructions show a higher productivity in varieties at higher developmental stages, whereas less advanced varieties rely more on entrenched verb-preposition strings (e.g. *refer to*) and WH + P combinations (e.g. *what...about*). At the same time, earlier Dynamic Model phases exhibit more variation in the preposition slot of pied-piped finite WH-relative clause constructions, which can be attributed to the weaker entrenchment of

constructions and a preference for explicit and transparent form-meaning pairings in these varieties. In sum, the results suggest that differences between the construction networks of varieties at different evolutionary stages result from a complex interplay of processing factors, input effects, second-language learning principles and, at least to a certain degree, cross-linguistic influence. In addition, the study shows that varieties of the same phase pattern very similarly in terms of their constructional choices, which provides support for Schneider's (2007) classification of varieties according to Dynamic Model stages.

Acknowledgements

I would like to express my gratitude to my supervisor Prof. Thomas Hoffmann for initiating this project, for his expertise, and for always believing in me. I am also grateful to Prof. Alexander Bergs for his support and for encouraging and helping me to pursue my interest in English linguistics in the first place. My colleague Dr. Thomas Brunner deserves many thanks for his valuable advice and ideas that greatly improved this study. Special thanks go to Prof. Peter Uhrig for his technical support regarding *The Red Hen Rapid Annotator*. Thanks are also due to my student assistants Jaqueline Hitér, Nicole Ratz, Tim Ullrich, Sarah Wehner, and, especially, Veronika Stampfer for their thorough work and their patience with my instructions. A heartfelt thank you goes to Bryan Mcleod and my Canadian family for never getting tired of my questions and for always being there for me. I would also like to thank Dr. Gerlinde Geiss-Mayer for her encouragement during the final stage of this project. Finally, none of this would have been possible without my friends and family. Thank you to my parents, Feli, Bruna and Sophia – I would not have survived these years without you!

Funding

This project was funded by a German Research Foundation (DFG) grant (HO 3904/7-1).

Contents

List of abbreviations.....	ix
List of figures	xi
List of tables	xiii
1. Introduction	1
1.1 Why preposition placement in World Englishes?	3
1.2 Theoretical framework	7
1.2.1 Usage-based Construction Grammar and New Englishes.....	7
1.2.2 A corpus-based approach	10
1.3 Research questions and outline	11
2. The evolution of PCEs from a cognitive perspective	13
2.1 World Englishes, New Englishes, and PCEs	13
2.2 The Dynamic Model from a Usage-based Construction Grammar Perspective	14
2.2.1 Models of World Englishes.....	14
2.2.2 Schneider's (2003, 2007) Dynamic Model	17
2.2.3 Hoffmann's (2021) Construction Grammar Approach to the Dynamic Model	21
2.2.3.1 Qualitative innovations from a constructionist perspective	21
2.2.3.2 Hoffmann's (2021) Dynamic Model Productivity hypothesis	24
2.3 Stylistic variation in PCEs from a cognitive perspective	28
2.4 Other relevant factors in the evolution of PCEs.....	33
2.4.1 Processing factors.....	33
2.4.1.1 Processing factors in World Englishes.....	34
2.4.1.2 Psycholinguistic evidence	42
2.4.2 Second-language learning principles	44
2.4.3 Cross-linguistic influence.....	47
2.4.3.1 Basic terms and assumptions.....	47
2.4.3.2 Cognitive approaches to cross-linguistic influence.....	51
2.5 Summary: The cognitive evolution of PCEs.....	53
3. Preposition placement from a usage-based perspective	55
3.1 The family of preposition placement constructions	55
3.1.1 The macro-constructional level.....	55

3.1.1.1	The pied-piping family	58
3.1.1.2	The stranding family	65
3.1.1.3	The preposition placement family	72
3.1.2	The meso-constructural level	77
3.2	Preposition placement and processing factors	81
3.2.1	The complexity of relative clause constructions.....	82
3.2.2	Type of PP	83
3.2.3	Type of phrase	90
3.2.4	Complexity of the FGD	92
3.2.5	Other processing-related factors	93
3.2.6	Summary: An overview of prototypical preposition placement constructions	96
3.3	Preposition placement constructions in World Englishes.....	97
3.3.1	The BrE vs. the KenE construction network	97
3.3.2	Quantitative differences between L1 and L2 Englishes	101
3.3.3	Qualitative innovations	103
3.4	Preposition placement in local languages.....	107
3.4.1	WH-interrogative constructions.....	108
3.4.2	Relative constructions.....	117
4.	Data and methods	126
4.1	The International Corpus of English.....	126
4.2	Data preparation and extraction.....	129
4.2.1	Corpus preparation.....	129
4.2.2	Data extraction.....	130
4.3	Data annotation.....	134
4.3.1	Linguistic factors	134
4.3.2	Extralinguistic factors	144
4.4	Statistical analysis.....	147
4.4.1	Mixed-effects logistic regression modeling.....	148
4.4.2	Hierarchical Configural Frequency Analysis (HCFA)	153
4.4.3	Large Number of Rare Events (LNRE) models.....	155
5.	Preposition placement in World Englishes.....	158
5.1	Hypotheses.....	158
5.2	Preposition stranding and pied-piping in World Englishes	160

5.2.1	Excluded tokens	160
5.2.2	Preposition placement across Dynamic Model stages	163
5.2.2.1	Model selection	163
5.2.2.2	Results	182
5.2.2.3	Summary	196
5.2.3	Preposition placement across varieties.....	197
5.2.3.1	Model selection	197
5.2.3.2	Results	202
5.2.3.3	Summary	212
5.2.4	Preposition placement in finite WH-relative clause constructions	214
5.2.4.1	Model selection	214
5.2.4.2	Results	221
5.2.4.3	Summary	231
5.3	Constructional fillers across varieties	233
5.3.1	V + P combinations	233
5.3.2	Fillers of the P + FILLER/FILLER + P slots in interrogatives.....	240
5.3.3	Fillers of the (ANTECEDENT +) P + WH slot in finite RCs.....	250
5.3.4	The P + WH slot in free RCs.....	264
5.3.5	Summary	266
5.4	The productivity of constructional slots.....	268
5.4.1	The V + P slot.....	268
5.4.2	The FILLER + P slot in free RCs and interrogatives.....	271
5.4.3	The ANTECEDENT and P + WH slots in finite RCs.....	273
5.4.4	Summary	277
5.5	Qualitative constructional innovations.....	278
5.6	Summary and discussion.....	292
6.	General conclusion	305
6.1	Main findings	305
6.2	Methodological implications.....	307
6.3	Avenues for future research	309
	Appendix	312
	References	349

List of abbreviations

AdjP	adjective phrase
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BNC	British National Corpus
BrE	British English
CanE	Canadian English
CG	Cognitive Grammar
EFL	English as a Foreign Language
ENL	English as a Native Language
ESL	English as a Second Language
FGD	filler-gap domain
finite RC	finite WH-relative
free RC	free relative
GloWbE	Corpus of Global Web-based English
GVIF	generalized variance inflation factor
HCFA	Hierarchical Configurational Frequency Analysis
HKE	Hong Kong English
ICE	International Corpus of English
ICE-CAN	ICE-Canada
ICE-EA	ICE-East Africa
ICE-EA-K	ICE-Kenya
ICE-EA-T	ICE-Tanzania
ICE-GB	ICE-Great Britain
ICE-HK	ICE-Hong Kong
ICE-IND	ICE-India
ICE-IRE	ICE-Ireland
ICE-JA	ICE-Jamaica
ICE-NIG	ICE-Nigeria
ICE-NZ	ICE-New Zealand
ICE-PHI	ICE-Philippines
ICE-SIN	ICE-Singapore
ICLE	International Corpus of Learner English
IndE	Indian English

IrE	Irish English
JamE	Jamaican English
KenE	Kenyan English
LDA	Latent Dirichlet Allocation
LNRE	Large Number of Rare Events
LRT	likelihood ratio test
ME	Middle English
NigE	Nigerian English
NP	noun phrase
NZE	New Zealand English
OE	Old English
PCEs	postcolonial Englishes
PGCH	Performance-Grammar Correspondence Hypothesis
PhilE	Philippine English
POS	part-of-speech
PP	prepositional phrase
SgE	Singapore English
TanE	Tanzanian English
VIF	variance inflation factor
VP	verb phrase

List of figures

2.1	The three ‘circles’ of English (Crystal 2003: 61).....	16
2.2	The unilateral relationship of the four sets of parameters of the Dynamic Model (Buschfeld et al. 2014: 6)	17
2.3	Linguistic innovations in the Dynamic Model from a constructionist perspective (based on Hoffmann 2021).....	22
2.4	The emergence of the [<i>discuss-about</i>]-construction (Gilquin 2017: 64)	46
2.5	Schematized version of Kellerman’s model of language transfer (adapted from Gass, Behney & Plonskey 2020: 146).....	50
3.1	The transitive verb-particle construction with its two allostructions (Cappelle 2006: 18)	57
3.2	Relationship between subject and object relative clauses and their construction neighbors (Diessel 2019: 208).....	63
3.3	The family of pied-piping constructions in BrE.....	64
3.4	The family of stranding constructions in BrE	71
3.5	The preposition placement family in BrE	76
3.6	Partial construction network of a formal pied-piped finite WH-relative clause construction	78
3.7	Preposition placement preference of individual PP types (based on Hoffmann 2011: 155, 172)	89
3.8	Magnitude estimation scores for stranding across different PP types in BrE (Hoffmann 2011: 182).....	99
3.9	Magnitude estimation scores for stranding across different PP types in KenE (Hoffmann 2011: 188).....	99
3.10	Usage-based construction network for BrE (Hoffmann 2011: 267)	100
3.11	Usage-based construction network for KenE (Hoffmann 2011: 268).....	100
5.1	Distribution of FILLER_TYPE across CLAUSE_TYPE	164
5.2	Distribution of CLAUSE_TYPE across VOICE	166
5.3	Distribution of topics across ICE categories	175
5.4	Topics per ICE variety in private dialogues.....	176
5.5	Topics per ICE variety in private correspondence (adapted from Mußemann 2024: 30)	177

5.6	Intercept adjustments of PREPOSITION and Wald confidence intervals (model: VARIETY_TYPE)	183
5.7	Relative filler frequencies per ICE variety	186
5.8	The effect of COMPLEXITY (logged) * CLAUSE_TYPE	187
5.9	The effect of CLAUSE_TYPE * FORMALITY	189
5.10	The effect of VARIETY_TYPE * FORMALITY	190
5.11	The effect of VARIETY_TYPE * CLAUSE_TYPE	190
5.12	The effect of VARIETY_TYPE * PP_TYPE	191
5.13	Distribution of CLAUSE_TYPE * FORMALITY across VARIETY	200
5.14	Intercept adjustments of PREPOSITION and Wald confidence intervals (model: VARIETY)	203
5.15	The effect of VARIETY * PP_TYPE	207
5.16	The effect of VARIETY * CLAUSE_TYPE	211
5.17	Distribution of FILLER_TYPE in the finite RC data compared to the finite RC, interrogative, and free RC data	216
5.18	Intercept adjustments of PREPOSITION and Wald confidence intervals (model: finite RCs)	221
5.19	The effect of COMPLEXITY (logged) * PP_TYPE	224
5.20	Boxplot: COMPLEXITY (logged) across VARIETY	225
5.21	The effect of RESTRICTIVENESS * VARIETY_TYPE	227
5.22	The effect of VARIETY_TYPE * TOPIC	229
5.23	Fillers of the V slot in WH... <i>about</i> free RC/interrogative constructions across variety phases	235
5.24	WH...V <i>like</i> constructions across variety phases	247
5.25	P + WH combinations in finite RCs across variety phases	254
5.26	<i>way/manner</i> -P- <i>which</i> across variety phases	259
5.27	The productivity of the V + P slot	269
5.28	The productivity of the V + P slot (informal/medium texts/everyday topics)	269
5.29	The productivity of the FILLER + P slot in interrogatives	272
5.30	The productivity of the FILLER + P slot in free RCs	272
5.31	The productivity of the ANTECEDENT slot in pied-piped finite RCs	274
5.32	The productivity of the P + FILLER slot in pied-piped finite RCs	275
5.33	Relative preposition frequencies per ICE variety	285
5.34	The preposition placement family of Phase V	298
5.35	The preposition placement family of Phase III	299

List of tables

1.1	Varieties investigated in this study along Schneider's (2007: 113-250) Dynamic Model	3
3.1	ICE-GB results for categorical clause contexts (Hoffmann 2011: 120)	69
3.2	Prototypical stranding and pied-piping constructions	96
3.3	Preposition placement in interrogative constructions in local languages.....	112
3.4	Preposition placement in relative constructions in local languages	119
4.1	ICE text categories (based on Nelson 2002: 4–5)	128
4.2	Stranded, pied-piped and non-standard tokens across varieties	134
4.3	Linguistic factors investigated in the present corpus study.....	135
4.4	Levels of PP_TYPE (Mußemann 2024: 11; based on Hoffmann 2011: 116).....	136
4.5	Interrater agreement: sample size and Kappa scores.....	144
4.6	Extralinguistic factors investigated in the present corpus study	145
4.7	The 20 most likely terms for the topics economy, government and technology	147
4.8	Example: HCFA output for VARIETY_TYPE * PREPOSITION * FILLER ..	154
5.1	Prototypical stranding and pied-piping constructions (repeated from 3.2.6).....	158
5.2	Distribution of idiosyncratic tokens across VARIETY.....	162
5.3	Distribution of non-standard tokens across VARIETY	163
5.4	Stranded and pied-piped <i>who</i> -tokens across varieties.....	165
5.5	Stranded and pied-piped <i>whom</i> -tokens across varieties.....	165
5.6	Distribution of stranding and pied-piping in simple, complex_zero and complex_that tokens across VARIETY_TYPE	167
5.7	HCFA results for VARIETY_TYPE/VARIETY * VERB_BRIDGE	167
5.8	Distribution of PP_TYPE across VARIETY_TYPE	169
5.9	Recoding of PP_TYPE.....	170
5.10	Distribution of stranding and pied-piping per CLAUSE_TYPE across VARIETY_TYPE	172
5.11	Distribution of PHRASE_TYPE across VARIETY_TYPE	172
5.12	Topics categorized according to formality (Mußemann 2024: 29).....	173
5.13	ICE text types categorized according to formality	175

5.14	Levels of FORMALITY	179
5.15	Distribution of FORMALITY across VARIETY_TYPE	179
5.16	Tokens with FILLER_LENGTH >= 3 across variety phases	180
5.17	Variance estimates and standard deviations of random effects (model: VARIETY_TYPE)	182
5.18	Coefficients of significant main effects (model: VARIETY_TYPE)	184
5.19	Coefficients of significant interaction effects (model: VARIETY_TYPE)	188
5.20	Stranded and pied-piped <i>be</i> -tokens across varieties	193
5.21	HCFA results for VARIETY_TYPE/VARIETY * PREPOSITION_PLACEMENT * CLAUSE_TYPE of <i>be</i> -tokens	194
5.22	Stranded and pied-piped respect, manner, degree PPs across varieties	197
5.23	Tokens with FILLER_LENGTH >= 3 across varieties	198
5.24	Distribution of PHRASE_TYPE across VARIETY	201
5.25	Variance and standard deviation estimates of random effects (model: VARIETY)	203
5.26	Coefficients of significant main effects (model: VARIETY)	204
5.27	Coefficients of significant interaction effects (model: VARIETY)	205
5.28	Stranding and pied-piping per PP type in WH-interrogatives across VARIETY_TYPE	209
5.29	Stranded and pied-piped finite RCs across varieties	215
5.30	Distribution of FILLER_TYPE in the finite RC data compared to the finite RC, interrogative, and free RC data	216
5.31	Tokens with FILLER_LENGTH > 1 across variety phases	217
5.32	Recoding of PP_TYPE (finite RCs)	217
5.33	Levels of TOPIC (Mußemann 2024: 14)	219
5.34	Distribution of TOPIC across VARIETY_TYPE	219
5.35	Variance estimates and standard deviations of random effects (model: finite RCs)	221
5.36	Coefficients of significant main effects (model: finite RCs)	223
5.37	Coefficients of significant interaction effects (model: finite RCs)	223
5.38	HCFA results for VARIETY_TYPE * V * P in free RCs/interrogatives	234
5.39	HCFA results for VARIETY * V * P in free RCs/interrogatives	236
5.40	WH... <i>talk-about</i> free RC/interrogative tokens per VARIETY	237
5.41	HCFA results for VARIETY_TYPE * V * P in finite RCs	238

5.42	Proportion of entrenched V + P combinations of the overall number of stranded finite RC tokens in Phases III and V	239
5.43	HCFA results for VARIETY * V * P in finite RCs	240
5.44	HCFA results for VARIETY_TYPE * P * FILLER in pied-piped interrogatives	243
5.45	HCFA results for VARIETY * P * FILLER in pied-piped interrogatives.....	244
5.46	HCFA results for VARIETY_TYPE * FILLER * P in interrogatives.....	245
5.47	<i>how...for</i> tokens per VARIETY_TYPE	246
5.48	HCFA results for VARIETY * FILLER * P in interrogatives.....	248
5.49	HCFA results for VARIETY_TYPE * P * WH in finite RCs	252
5.50	HCFA results for VARIETY * P * WH in finite RCs	253
5.51	HCFA results for VARIETY_TYPE * ANTECEDENT * P_WH in finite RCs	257
5.52	<i>way/manner-P-which</i> across variety phases	260
5.53	HCFA results for VARIETY * ANTECEDENT * P_WH in finite RCs.....	263
5.54	Pied-piped free RC tokens across varieties	264
5.55	P + WH combinations in free RCs	265
5.56	χ^2 - and <i>p</i> -values of the LNRE models of the V + P slot.....	268
5.57	χ^2 - and <i>p</i> -values of the LNRE models of the V + P slot (informal/medium texts/everyday topics).....	270
5.58	χ^2 - and <i>p</i> -values of the LNRE models of the FILLER + P slot in interrogatives	271
5.59	χ^2 - and <i>p</i> -values of the LNRE models of the FILLER + P slot in free RCs	271
5.60	χ^2 - and <i>p</i> -values of the LNRE models of the ANTECEDENT slot in pied-piped finite RCs.....	273
5.61	χ^2 - and <i>p</i> -values of the LNRE models of the P + FILLER slot in pied-piped finite RCs	275
5.62	Recoding of CATEGORY into “less/more spontaneous”	279
5.63	Non-standard preposition placement types across VARIETY	280
5.64	HCFA results for VARIETY * NONSTANDARD_TYPE * CLAUSE_TYPE * PP_TYPE * CATEGORY	281
5.65	HCFA results for VARIETY/VARIETY_TYPE * V of missing tokens with lexicalized/complement_like PPs.....	282
5.66	Prepositions in unexpected tokens	285
5.67	P + WH combinations in mismatching tokens of Phases III and IV	288

5.68 P + WH combinations in extra tokens	289
5.69 Summary of study results (Hypotheses I - VII).....	292

1. Introduction

Speakers can often choose between two or more constructions with a very similar meaning. While the first variationist studies simply sought to relate these “alternate ways of saying ‘the same’ thing” (Labov 1972: 188) to social variables (Eckert 2012), the present study adopts a cognitive perspective and aims to provide more insights into the representation of competing constructions in the mental grammars of speakers of varieties of English worldwide. Some of the most well-known alternation phenomena in English that have been analyzed within the field of cognitive linguistics include the so-called “genitive alternation” (*John’s book* vs. *the book of John*; see e.g. Hinrichs & Szmrecsanyi 2007; Gries & Wulff 2013; Jankowski & Tagliamonte 2014; Szmrecsanyi et al. 2016, 2017; Heller, Szmrecsanyi & Grafmiller 2017), the “dative alternation” (*John gave Mary the book* vs. *John gave the book to Mary*; see e.g. Bresnan et al. 2007; Bresnan & Ford 2010; Bernaisch, Gries & Mukherjee 2014; Szmrecsanyi et al. 2016, 2017; Röthlisberger, Grafmiller & Szmrecsanyi 2017), “particle placement” (*John **picked up** the book* vs. *John **picked** the book **up***; see e.g. Gries 2003; Lohse, Hawkins & Wasow 2004; Cappelle 2009; Gilquin 2015; Szmrecsanyi et al. 2016; Grafmiller & Szmrecsanyi 2018; Paquot, Grafmiller & Szmrecsanyi 2019; Wulff & Gries 2019) and “that-complementation” (*I thought you would win* vs. *I thought **that** you would win*; see e.g. Tagliamonte & Smith 2005; Cacoullos & Walker 2009; Jaeger 2010; Wulff, Lester & Martinez-Garcia 2014; Wulff, Gries & Lester 2018). The phonological, morphosyntactic, semantic, discourse-functional and processing-related variables that play a role in these alternations have been explored extensively not only in Standard varieties but also with respect to learner Englishes and postcolonial Englishes (PCEs; cf. 2.4.1.1). Another alternation phenomenon which has also been studied widely in Standard Englishes but which has so far received comparatively little attention in the field of World Englishes is preposition placement.

The term “preposition placement” (Hoffmann 2011) refers to the alternation between preposition stranding and preposition pied-piping in English. As illustrated in the relative clause example in (1.1), Standard Englishes have two options of placing a preposition: the preposition can either remain in-situ without its complement, i.e. it can be “stranded” (1.1a), or it can appear in front of the WH-word (1.1b), which is referred to as “pied-piping” (Hornstein & Weinberg 1981; Ross 1986; Van den Eynden 1996; Pullum & Huddleston 2002; Hoffmann 2006, 2007a, 2011, 2013).

- (1.1) (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>

Stranded and pied-piped prepositions occur in so-called “filler-gap” constructions such as relative clauses, WH-interrogatives, WH-exclamatives, *it*-clefts and topicalizations (cf. 3.1), which are very difficult to process (Pollard & Sag 1994: 157; Sag 2010; Hoffmann 2022: 229–234). All filler-gap constructions have in common that they consist of a “filler” (e.g. *which* in (1.1a), *in which* in (1.1b)) and a “gap”, i.e. the position in which the phrase represented by the filler would appear in a declarative sentence (e.g. *We live **in the real world*** in (1.1a)). Consequently, when parsing a filler-gap dependency, the human processor has to hold the filler in working memory while identifying the corresponding gap. This process is difficult in itself as the gap is not indicated by an explicit element but can only be inferred from context. At the same time, the material standing in the path from the filler to the gap also has to be processed, which leads to an additional increase in processing load (Hawkins 1999: 246–247, 2004: 173). Hence, it is not surprising that previous studies have shown that the competition between preposition stranding and pied-piping is largely determined by two factors, namely processing constraints and stylistic variables (Gries 2002; Hoffmann 2011: 93–98). As these aspects are also pressing issues in the study of PCEs (cf. 2.4), the present analysis offers a comprehensive investigation of preposition placement in World Englishes in order to shed more light on the factors that influence the constructional choices of speakers of varieties of English worldwide.

In doing so, the study draws on data from the International Corpus of English (ICE; Greenbaum & Nelson 1996) and explores the constructional competition between preposition stranding and pied-piping in 12 varieties of English at different stages of Schneider’s Dynamic Model with a wide range of typologically very distinct L1s. As shown in table 1.1, the analysis includes the Phase-V varieties British English (BrE), Canadian English (CanE), Irish English (IrE) and New Zealand English (NZE), the Phase-IV varieties Jamaican English (JamE) and Singapore English (SgE) as well as the Phase-III varieties Kenyan English (KenE), Tanzanian English (TanE), Nigerian English (NigE), Indian English (IndE), Philippine English (PhilE), and Hong Kong English (HKE). Following recent cognitive studies in World Englishes research (Hoffmann 2011, 2014, 2019, 2020a, 2021; Brunner & Hoffmann 2020; Brunner 2021; Hoffmann & Brunner 2022; Horsch 2023), the analysis adopts a combination of Usage-based Construction Grammar and New Englishes approaches in order to contribute to the understanding of the cognitive principles underlying the evolution of New Englishes.

This introduction paves the way for this endeavor by clarifying the motivation behind the study (1.1) and by establishing the theoretical and methodological framework chosen for the analysis (1.2). Moreover, it will give an overview of the main research questions and the outline of the study (1.3).

Table 1.1: Varieties investigated in this study along Schneider's (2007: 113-250) Dynamic Model

Phase III	Phase IV	Phase V
Kenyan English (KenE)	Jamaican English (JamE)	British English (BrE)
Tanzanian English (TanE)	Singapore English (SgE)	Irish English (IrE)
Nigerian English (NigE)		Canadian English (CanE)
Indian English (IndE)		New Zealand English (NZE)
Philippine English (Phile)		
Hong Kong English (HKE)		

1.1 Why preposition placement in World Englishes?

The most comprehensive study on preposition placement in World Englishes that has been conducted to date is Hoffmann's (2011) investigation of preposition placement in BrE and KenE. While Hoffmann (2011: 278) found that the constructional choices of both BrE and KenE speakers are strongly driven by processing factors, he also identified two major differences between the varieties: First, the results of a series of magnitude estimation experiments indicated that L2 KenE speakers rate stranding with adjunct prepositional phrases (PP) and pied-piping with prepositional verbs lower than L1 BrE speakers (Hoffmann 2011: 187–188). This clearly reflects a processing effect. In line with Hawkins' (1999: 277) "Avoid Competing Subcategorizers" principle, which claims that „[t]he human processor prefers to avoid garden paths that result from competing subcategorizers within [a filler-gap domain]", pied-piping is generally easier to process than stranding as pied-piped prepositions facilitate the integration of the filler so that garden paths are avoided. Since the human processor always aims to identify the earliest possible gap site (Hawkins 1999: 247), in the stranded example (1.2a) it is likely that the filler *what* is mistakenly identified as the object of *wear*. In contrast, the pied-piped preposition in (1.2b) avoids such a misanalysis (Hawkins 1999: 277). However, if the preposition and the verb are strongly associated with each other (i.e. e.g. with prepositional verbs), stranding can offer a processing advantage over pied-piping (Hawkins 1999: 260, fn. 15; Pullum & Huddleston 2002: 629; Hoffmann 2011: 59).

Hence, example (1.3a), in which the prepositional verb *refer to* is not broken up, is preferred over (1.3b).

- (1.2) (a) [*what*]_i *did you* [*wear* (O_i) *it with* O_i] <ICE-IRE:S1A-063\$B>
 (b) [*with what*]_i *did you* [*wear it* O_i]
- (1.3) (a) *I'm not quite sure exactly what the member is referring to* <ICE-CAN:S1B-057#115:1:A>
 (b) *And uh members of the Committee will know to what I refer* <ICE-IRE:S2A-068\$A>

Since both pied-piping with adjunct PPs and stranding with prepositional verbs should thus be preferred from a processing perspective, Hoffmann (2011: 270) concluded that “the Kenyan speakers, who can be assumed to have received less input, tend to favour these prototypical realizations of preposition placement more than the British speakers”. Hoffmann’s (2011) results thus indicate that the constructional choices of L2 speakers are more strongly affected by processing constraints than those of L1 speakers.

The second factor that led to differences between the BrE and the KenE construction networks in Hoffmann’s (2011) study is formality, which turned out to have a greater effect on preposition placement in BrE than in KenE. While BrE speakers clearly favor pied-piping in formal and stranding in informal relative clauses, KenE speakers exhibit only a very weak stylistic effect and favor pied-piping regardless of the context (Hoffmann 2011: 155, 167). Hoffmann’s (2011) study thus demonstrated that differences between the construction networks of L1 and L2 speakers result from an interaction of processing factors and stylistic variables.

However, despite Hoffmann’s (2011) promising results, so far, no comprehensive study on preposition stranding and pied-piping has been conducted in order to assess the validity of his claims in a wider range of L1 and L2 varieties. Support for the prediction that processing factors play a more important role in L2 Englishes comes from Hoffmann’s (2013: 114–116) experimental study, in which L2 German speakers also rated pied-piping with prepositional verbs lower than with time or location adjunct PPs, while the ratings of L1 speakers were equally high for both groups. Likewise, in an experimental study conducted by Jach (2018: 286), L2 German and L2 Chinese speakers rated pied-piping with prepositional verbs lower than L1 English speakers. Importantly, however, all these studies that point to processing-related differences between varieties are based on experimental data and should thus also be corroborated by corpus-based research (Hoffmann 2006: 167). Furthermore, with respect to other alternation

phenomena, research on the role of processing factors in L1 and L2 Englishes has provided conflicting results (cf. 2.4.1.1). While studies on *that*-complementation (e.g. Wulff, Lester & Martinez-Garcia 2014; Wulff, Gries & Lester 2018) or particle placement (e.g. Gilquin 2015: 64–65; Wulff & Gries 2019) suggested that the constructional choices of L2 speakers are, at least in parts, more strongly driven by processing constraints than those of L1 speakers, research on the dative (Bernaisch, Gries & Mukherjee 2014; Röthlisberger, Grafmiller & Szmrecsanyi 2017) and the genitive alternation (Gries & Wulff 2013; Heller, Szmrecsanyi & Grafmiller 2017) found that processing constraints are at work in a very similar way in L1 and L2 Englishes. Hence, for instance, Bernaisch, Gries & Mukherjee (2014: 27) concluded, based on their analysis of the dative alternation in South Asian Englishes, “that factors determining processing and, thus, ultimately constructional choices are widely applicable to all varieties of English and may therefore represent features shared by all Englishes”. Although recent years have seen the emergence of a number of studies investigating the effect of processing factors on speakers of L1 and L2 Englishes, this shows that there are still many unknown aspects about this matter.

Hoffmann’s (2011) finding of a weaker formality effect in KenE also needs to be corroborated by a thorough investigation of preposition placement that takes into account a wide range of text types across various L1 and L2 Englishes. Most studies on preposition placement in World Englishes only investigated very low token numbers from individual varieties, and were not based on multifactorial statistical analyses. Suárez-Gómez (2014, 2015) looked at preposition placement in informal, spoken language in Southeast Asian Englishes, but only analyzed 40 relative clause tokens from IndE, 57 from HKE, and 31 from SingE (Suárez-Gómez 2014: 260). Dayag (2016) investigated 72 instances of preposition stranding and pied-piping in PhilE, but mostly focused on qualitative aspects and did not distinguish between relative and interrogative tokens although clause type is known to strongly affect the choice between stranding and pied-piping (cf. 3.1). Hence, both Suárez-Gómez’ (2014: 260) finding that IndE has a stronger pied-piping preference than HKE and SgE, which is in accordance with the widespread assumption that IndE tends to prefer formal variants, and Dayag’s (2016: 118) claim that the strong pied-piping preference of PhilE in both spoken and written registers reflects the “stylistic under-differentiation” (Gonzalez 1991: 334) of PhilE, have to be taken with a grain of salt and need to be corroborated by studies using larger datasets. Likewise, Adejare’s (2021) investigation of ICE-Nigeria, which observed that pied-piping is slightly more frequent in academic writing than in popular writing, exhibits serious

problems since Adejare (2021: 52) apparently also included phrasal verbs such as *turn on/off* (as in *turn the light on/off*) in her analysis of preposition placement. Jach's (2021) analysis of the French, German, Italian, Spanish, Chinese, and Japanese components of the International Corpus of Learner English (ICLE), which suggested that L1 transfer leads to a pied-piping preference in European L2 Englishes (Jach 2021: 362), excluded informal text types altogether and only focused on argumentative essays (Jach 2021: 347). At the same time, Hoffmann's (2011) finding of a weaker formality effect in KenE is in accordance with studies on other linguistic phenomena (cf. 2.3) which also indicated that PCEs tend to be characterized by "a lack of register distinctions" (Mußemann 2024: 2) as L2 speakers frequently overuse variants typical of written language in spoken registers, while their writing tends to be characterized by features typical of spoken discourse (Deshors, Götz & Laporte 2018: 13–14). By offering a detailed, corpus-based analysis of preposition placement in 12 varieties of English, the present study thus seeks to contribute to this scholarship by providing new insights into the ways in which processing factors, stylistic variables, and cross-linguistic influence affect the constructional choices of speakers of varieties of English worldwide.

Furthermore, preposition placement is an ideal phenomenon to explore the process of "structural nativization" (Schneider 2007: 5), i.e. the emergence of locally distinctive constructions in PCEs, which typically starts out "at the interface between grammar and lexis" (Schneider 2007: 46). Since prepositions frequently appear in constructions with non-compositional meanings (such as prepositional verbs), it is not surprising that many studies on structural nativization in PCEs report innovative uses of prepositions such as a tendency to replace idiomatic prepositions with more transparent variants (e.g. the use of *critical against* instead of *critical to*; see e.g. Nesselhauf 2009; Schneider & Gilquin 2018: 57–58). In the field of World Englishes, prepositions have therefore also been referred to as "mutating species" (Gilquin & Granger 2011: 60). With respect to preposition placement, Hoffmann (2011: 133–134) found that speakers of KenE do not only use structures with unexpected prepositions such as (1.4), in which *conditions to which* is used instead of the Standard variant *conditions under/in which*, but also constructions with doubled (e.g. *in* in 1.5) or missing (e.g. *for* in (1.6)) prepositions.

(1.4) *Hostels no longer provide **conditions to which** students can study efficiently* <ICE-EA-K:S2B032K>

(1.5) *On inspection I found they were copper wires which we use with our telecommunications in state **in** which they were **in** they could not be used as they were coiled and burnt* <ICE-EA-K:S1B067K>

(1.6) *How much do they sell it* <ICE-EA-K:S1A012K>

These non-standard types of preposition placement should be analyzed in a greater number of PCEs in order to investigate to what extent these innovations can be attributed to processing constraints and second-language learning effects (see e.g. Mesthrie & Bhatt 2008: 174; Schneider & Gilquin 2018). The present analysis of preposition placement thus also aims to shed more light on the cognitive processes that lead to qualitative constructional innovations in PCEs.

1.2 Theoretical framework

1.2.1 *Usage-based Construction Grammar and New Englishes*

The theoretical framework chosen for the present study is a combination of Usage-based Construction Grammar and New Englishes approaches. Construction Grammar (e.g. Croft 2001; Goldberg 2006, 2019; Diessel 2019; Hilpert 2019; Hoffmann 2022), a theory that has recently received strong support from psycho- (see e.g. Wardlow Lane & Ferreira 2010; Bencini 2013; Perek 2015) and neurolinguistic research (see e.g. Cappelle, Shtyrov & Pulvermüller 2010; Pulvermüller, Cappelle & Shtyrov 2013) as well as studies on L1 (see e.g. Dąbrowska 2000; Tomasello 2003; Diessel 2004, 2013) and L2 acquisition (see e.g. Ellis 2003, 2013; Gries & Wulff 2005, 2009), maintains that speakers store all linguistic knowledge in the form of constructions, i.e. as “arbitrary pairing[s] of (phonological/syntactic) FORM and MEANING” (Hoffmann 2022: 4; also see Croft & Cruse 2004: 255; Goldberg 2006: 5). While Mainstream Generative Grammar (Chomsky 1965, 1995) and traditional approaches to grammar (e.g. Quirk et al. 1985) maintain a clear-cut distinction between the syntax and the lexicon, constructionists assume a continuum ranging from fully abstract constructions such as the resultative construction ([X V Y Z] – ‘X causes Y to become Z by V-ing’), which contains various slots that can be filled freely, to substantive constructions which are phonologically fully fixed (e.g. green [gri:n] – ‘green’; Croft & Cruse 2004: 255; Hoffmann & Trousdale 2013a: 1; Hoffmann 2022: 19). Following Traugot (2008: 32), the most schematic constructions are also referred to as “macro-constructions”, while the most substantive constructions are known as “micro-constructions”. Furthermore, there are so-called “meso-constructions” (Traugot 2008: 32), i.e. partly schematic constructions with both substantive and schematic slots, such as the *X take Y for granted* construction ([X TAKE Y fə ɡrɑ:ntɪd] – ‘X doesn’t value Y’; Goldberg 2006: 5; Croft & Cruse 2004: 255–256; Hoffmann &

Trousdale 2013a: 2). All these constructions are stored in a network commonly referred to as the “constructicon” (Jurafsky 1992: 302; Goldberg 2003: 219; also see Goldberg 2019; Diessel 2019). At the same time, the term “construct” is used to refer to concrete utterances in language use, which frequently result from the combination of several constructions (Hilpert 2019: 12).

While Construction Grammar subsumes various approaches that differ in their views on the entrenchment of substantive and schematic constructions (for an overview, see e.g. Hoffmann & Trousdale 2013b), the present study is situated within the framework of Usage-based Construction Grammar, which emphasizes the importance of authentic language input for the emergence of cognitive constructional representations. Assuming that the entrenchment of constructions is tied to domain-general cognitive principles such as categorization, chunking or rich memory storage (Bybee 2010: 7–9), usage-based approaches maintain that speakers do not only store form-meaning pairings that are “not strictly predictable from [their] component parts or from other previously established constructions” (Goldberg 1995: 4) but any pattern that “occur[s] with sufficient frequency” (Goldberg 2006: 5). In other words, if a speaker is exposed to e.g. different instances of a stranded *WH...about* construction such as *What does Sarah talk about?*, *What did you laugh about?* or *What did Bob speak about?*, over time, a partly schematic *WH...about* interrogative construction may become entrenched in the speaker’s constructicon. As pointed out by Schmid (2020: 43), cognitive representations of language and language use thus continually reinforce each other: “[r]epeated processing leads to entrenchment; entrenchment in turn is conducive to repeated activation”. As a result, deeply entrenched constructions are generally associated with a high degree of automaticity and fluency. In his concept of Cognitive Grammar (CG), which shares many views with Construction Grammar (Hoffmann & Trousdale 2013a: 7), Langacker (2008: 16) explains the relationship between entrenchment and automatization as follows:

Automatization is the process observed in learning to tie a shoe or recite the alphabet: through repetition or rehearsal, a complex structure is thoroughly mastered, to the point that using it is virtually automatic and requires little conscious monitoring. In CG parlance, a structure undergoes progressive **entrenchment** and eventually becomes established as a **unit**. (Langacker 2008: 16; original emphasis)

Consequently, especially deeply entrenched micro- or meso-constructions at a low level of schematicity have a high chance of being activated in spontaneous language use (Langacker 1999: 104–107; Schmid 2020: 214). Importantly, however, an increase in input frequency does not always automatically go along with the entrenchment of abstract constructions (Croft 2001; Barðal 2008, 2011). Only if speakers encounter a construction with a high type frequency, can they make bottom-up generalizations over more

substantive constructions, which may ultimately lead to the entrenchment of an abstract constructional template. At the same time, substantive constructions can become entrenched as a result of a high token frequency (Croft & Cruse 2004: 292–293; Goldberg 2006: 98–101; Bybee 2010: 9, 95–96). The construction thus exhibits redundancy as speakers store constructions at varying degrees of schematicity (Hilpert 2013: 201–203), i.e. both specific, substantive instances of constructions as well as more abstract constructional representations that emerge as a result of bottom-up generalizations (Goldberg 2006: 55; Bybee 2010: 15).

With respect to the present study, Usage-based Construction Grammar is particularly advantageous as it emphasizes the importance of taking into account preposition placement constructions at all levels of abstraction. Hoffmann (2011: 160–165) showed that speakers do not only store abstract preposition placement constructions such as a formal pied-piped WH-relative construction but also meso-constructions such as a partly schematic *way-in-which* construction (1.7), in which the string *way in which* is fully specified. By adopting a constructionist framework, the present study thus seeks to identify cross-varietal differences in the use and entrenchment of preposition placement constructions at all levels of schematicity.

(1.7) *And I congratulate him also on the **way in which** he introduced this measure* <ICE-GB:S1B-051 #4:1:A>

The second theoretical framework chosen for the present study is Schneider's (2003, 2007) Dynamic Model of the evolution of PCEs, which assumes that the emergence of New Englishes can be described by a uniform evolutionary process consisting of up to five developmental stages. According to Schneider (2003, 2007), over time, the colonists and the indigenous population increasingly become aware of a common identity, which contributes to the spread of English and, ultimately, also to linguistic change. Schneider's model provides a useful theoretical foundation for the present endeavor as it accounts for the fact that different stages of evolution are associated with differences in exposure to input, which can be expected to lead to differences between variety phases with respect to the effect of processing constraints and stylistic variables (cf. chapter 2). Moreover, Hoffmann (2021) complemented Schneider's sociolinguistic model with a Usage-based Construction Grammar perspective (cf. 2.2.3), arguing that the productivity of constructional slots should be correlated with the evolutionary stage of a variety. More specifically, Hoffmann (2021: 21) predicts that, due to differences in the input frequency of constructions, varieties at higher Dynamic model

stages should exhibit a higher productivity of constructional slots than varieties at lower stages of evolution. In addition, Hoffmann (2021: 19) provides a constructionist explanation of structural innovation in PCEs, which can be said to typically emerge at the meso-constructional level. By combining Usage-based Construction Grammar and New Englishes approaches, the present study seeks to test these predictions in order to provide a comprehensive account of the construction networks of different variety phases, which will allow more insights into the cognitive aspects that shape the emergence of PCEs.

1.2.2 *A corpus-based approach*

This investigation is a corpus-based study which will use three different statistical methods to identify entrenched stranding and pied-piping constructions at different levels of schematicity in 12 varieties of English (for details regarding the statistical techniques, see section 4.4). While mixed-effects logistic regression modeling (Baayen 2008: 241–302) will be applied in order to reveal differences between varieties with respect to processing constraints and stylistic variables, Hierarchical Configural Frequency Analysis (HCFA) will be used to identify significant associations between individual varieties/variety types and specific constructional fillers (Gries 2009: 240–252; Brunner & Hoffmann 2020; Hoffmann 2021). Moreover, the study will draw on “Large Number of Rare Events” (LNRE) models to compare the productivity of constructional slots across variety phases (Zeldes 2013; Baroni & Evert 2014; Brunner & Hoffmann 2020; Hoffmann 2020a, 2021). It is important to note, though, that results from corpus studies are always based on aggregated data from several speakers. Entrenchment being a process that takes place in the mental grammars of individual speakers, the results of corpus studies can thus never reflect direct cognitive entrenchment (Blumenthal-Dramé 2012; Stefanowitsch & Flach 2017: 122; Schmid 2020: 217). Instead, Schmid (2010: 117) suggests that statistically significant effects obtained in corpus-based studies should be taken as indicators for the “degree of *conventionalization*” of constructions. In addition to this “*corpus-as-output* view” (Stefanowitsch & Flach 2017: 102), the results can also be interpreted from a “*corpus-as-input* view” (Stefanowitsch & Flach 2017: 103). Put differently, from a usage-based perspective, corpus data can be taken as “a (more or less representative) sample of the language use that members of a particular speech community are exposed to during language acquisition and on the basis of which they construct their mental representations in the first place” (Stefanowitsch & Flach 2017: 103).

1.3 Research questions and outline

In the light of the above, the study aims to answer the following research questions:

- R1. Which factors influence the choice between preposition stranding and pied-piping in Englishes around the world?
- R2. Do processing factors have a stronger effect in L2 Englishes?
- R3. Is there a correlation between the strength of stylistic effects and Dynamic Model stages?
- R4. Does the evolutionary state of the New English varieties correlate with the productivity of preposition stranding and pied-piping constructions?
- R5. Can the use of non-standard preposition placement constructions be linked to processing factors and general principles of L2 learning?
- R6. Does cross-linguistic influence lead to quantitative differences between varieties with respect to the use of preposition placement constructions?

The first part of the study (chapters 2-3) will provide the relevant theoretical background that is necessary to investigate these research questions. After introducing some basic terms (2.1), chapter 2 will explain the evolution of PCEs from a cognitive perspective, with a particular focus on Hoffmann's (2021) Construction Grammar Approach to the Dynamic Model (2.2). Furthermore, 2.3 will provide a cognitive explanation of stylistic variation in New Englishes, while 2.4. will outline other factors which are known to play a role in the evolution of PCEs and which should be taken into account in this analysis.

Chapter 3 will deal with the linguistic phenomenon investigated in the present study, namely preposition placement in filler-gap constructions. In 3.1, it will be argued that the preposition stranding and pied-piping constructions that exist in English can be described as a family of interconnected stranding and pied-piping constructions. 3.2 will outline the relationship between preposition placement and processing factors, while the third part of this chapter will give an overview of previous studies on preposition stranding and pied-piping in World Englishes that are relevant for the present investigation (3.3). Finally, 3.4 will focus on the preposition placement constructions that exist in the local languages of the PCEs at hand.

Chapter 4 will provide the methodological basis for the corpus study. After a short description of the ICE corpora (4.1), the data extraction (4.2) and the data annotation process (4.3) will be described in detail. Moreover, this chapter will introduce the three statistical techniques that will be used to analyze the corpus data (4.3).

Chapter 5 will provide the main hypotheses (5.1) and the results of the corpus-based analysis. Section 5.2 will present the findings of three generalized linear mixed-effects models, which suggest that less advanced varieties are more strongly affected by processing factors than varieties at higher stages of evolution, whereas variables related to formality and topic have a weaker effect in earlier Dynamic Model phases. In addition, section 5.3 will identify significant associations between fillers of constructional slots and varieties/variety types, while 5.4 will compare the productivity of constructional slots across variety phases. Preposition stranding constructions turn out to be more productive in varieties at higher Dynamic Model stages, while speakers of less advanced varieties rely more on entrenched constructional fillers particularly in non-prototypical stranding contexts and in more informal usage situations. At the same time, the least advanced varieties exhibit the highest productivity of the P + WH slot (e.g. *in which*) in pied-piped finite WH-relative clause constructions. The last part of the corpus study focuses on qualitative constructional innovations (5.5), while a summary and discussion of the findings is given in 5.6. Finally, chapter 6 summarizes the overall study results.

2. The evolution of PCEs from a cognitive perspective

As outlined in 1.2.1, the present study adopts a combination of Usage-based Construction Grammar and New Englishes approaches to analyze the use of preposition stranding and pied-piping constructions in 12 varieties of English. This chapter seeks to lay the foundations for the corpus-based analysis presented in chapter 5 by explaining the evolution of PCEs from a cognitive perspective. While 2.1 will introduce some basic terms, section 2.2 will focus on Schneider's (2003, 2007) Dynamic Model and, in particular, on Hoffmann's (2021) Construction Grammar Approach to Schneider's sociolinguistic model. 2.3 will provide a cognitive explanation of stylistic variation in PCEs, while 2.4 will discuss other factors that play a role in the evolution of PCEs, viz. processing factors, second-language learning principles, and cross-linguistic influence. The main aspects of the chapter will be summarized in 2.5.

2.1 World Englishes, New Englishes, and PCEs

In the literature, several terms are used to refer to varieties of English. The three most common labels, which will also be used in the present study, are "World Englishes", "New Englishes", and "Postcolonial Englishes (PCEs)". These terms can be defined as follows:

The label "World Englishes", which was introduced by Kachru to draw attention to postcolonial varieties of English (Schneider 2007: 3), refers to both L1 and L2 Englishes spoken all around the world, regardless of the history, the evolutionary background, the norm orientation or the status of a variety. Consequently, this term also covers Standard varieties such as BrE (Schneider 2020: 31).

The term "New Englishes" was introduced by Pride (1982) and Platt, Weber & Ho (1984) to describe L2 varieties of English which have emerged as a result of colonialism and which are usually spoken in addition to local languages. These varieties are typically associated with formal domains of use such as politics, education or the media. On the linguistic level, they are characterized by phonological, lexical, grammatical and pragmatic innovations (Platt, Weber & Ho 1984: 2–3). However, the term has been criticized as it has been argued "that the label *new* reflected primarily a shift of attention in western, Anglocentric scholarship" (Schneider 2007: 3; original emphasis).

The term “Postcolonial Englishes” (Schneider 2007) is more neutral and objective than the label New Englishes. Moreover, it explicitly emphasizes that the varieties under discussion “are products of a specific evolutionary process tied directly to their colonial and postcolonial history” (Schneider 2007: 3). Note that L1 varieties, such as NZE or American English, also have a history of colonialism and would thus normally also be considered PCEs (Schneider 2020: 31). However, in order to distinguish these classic L1 varieties from the more prototypical PCEs, such as IndE or KenE, in which English is spoken as an L2 alongside local languages, in this study, the term “Postcolonial Englishes (PCEs)” will only be used to refer to L2 varieties that emerged as a result of a colonial contact setting.

In the present analysis, the labels “New Englishes” and “PCEs” thus refer to varieties assigned to Phases III and IV in Schneider’s (2007) classification of varieties according to Dynamic Model stages (cf. table 1.1), while the term “World Englishes” covers all varieties regardless of their evolutionary stage.

2.2 The Dynamic Model from a Usage-based Construction Grammar Perspective

This study adopts Schneider’s (2003, 2007) Dynamic Model of the evolution of PCEs. However, in order to demonstrate the superiority of Schneider’s sociolinguistic approach over other models, section 2.2.1 will provide a brief overview of models of World Englishes.

2.2.1 Models of World Englishes

Over the course of time, various models have been developed in order to categorize varieties of English. In one of the first attempts, Quirk (1985) developed a concept originally made by Barbara Strang into the widely known distinction between ENL (English as a Native Language), ESL (English as a Second Language) and EFL (English as a Foreign Language) countries (Schneider 2020: 32). While English normally serves as the mother tongue in ENL countries such as Great Britain or Ireland, in EFL countries such as Spain or Senegal, it is exclusively acquired as a foreign language and does not have any internal functions. ESL countries such as India can be situated somewhere in-between. In these countries, speakers typically acquire other languages as their first language, but English follows as a second language and fulfils important functions in e.g.

politics or education (Quirk 1985: 1–2). However, despite its clear-cut conceptualization, the ENL-ESL-EFL classification has been criticized as being “dry and unappealing” (Schneider 2020: 32). Furthermore, it is almost exclusively based on the concept of mother tongue and thus clearly centered on ENL varieties (Saraceni 2010: 40).

In contrast, Kachru’s (1985) “Three Circles” model, which aims to emphasize the value of linguistic diversity, takes a more pluralistic approach (Buschfeld & Kautzsch 2020: 55). The Three Circles model, in which ENL, ESL and EFL varieties correspond to three concentric circles (cf. figure 2.1), namely the “Inner Circle”, the “Outer Circle” and the “Expanding Circle” respectively, adds a sociopolitical and historical dimension to Quirk’s ENL-ESL-EFL classification. While Inner Circle varieties, such as Great Britain, Ireland, Canada or New Zealand, which Kachru (1985: 12) refers to as “the traditional bases of English”, are considered “norm-providing” (Kachru 1985: 16), Outer Circle varieties, which correspond to the notion of New Englishes, i.e. varieties in which English has “an extended functional *range* in a variety of social, educational, administrative, and literary domains” (Kachru 1985: 13; original emphasis), are “norm-developing” (Kachru 1985: 17). The EFL varieties found in the Expanding Circle, such as China or Russia, are “norm dependent” and thus exonormatively oriented toward the norms set by the Inner Circle varieties (Kachru 1985: 17; Mesthrie & Bhatt 2008: 29). Compared to previous approaches, Kachru’s model draws more attention to Outer Circle varieties, thereby challenging the conceptualization of ENL countries as being superior that had shaped earlier models (Buschfeld & Kautzsch 2020: 55). Moreover, according to Schneider (2003: 237), the Three Circles model “emphasizes that the English language belongs to all who use it, and that the most vigorous expansions and developments of the language can be observed in Outer and Expanding Circle countries”.

Although Kachru adopted a more pluralistic view than previous scholars, the Three Circles model has also received its share of criticism. As Kachru’s nation-based approach implies homogenous varieties that correspond directly to countries, “it ignores certain facets of complex realities” (Schneider 2007: 12) that characterize the linguistic landscape all around the world. For instance, Kachru’s model is unable to account for ESL and EFL varieties that exist in typical ENL countries, such as French Canadian English in Canada or Native American varieties of English in the US (Schneider 2007: 12). Likewise, it is not compatible with the complex linguistic situation in South Africa, which accommodates several ENL and ESL varieties of English such as White South African English, Black South African English or Indian South African English, and can thus not be clearly assigned to one of the three circles (Schneider 2007: 13, 173–188).

Furthermore, Kachru's classification lacks a diachronic dimension (Buschfeld & Kautzsch 2020: 63). As Schneider pointed out (2020: 32), due to its static nature, the Three Circles model "fail[s] to reflect the vigorous spread of English and changes of its status in many regions over the last few decades". A more realistic model should thus treat the Inner Circle/Outer Circle and Outer Circle/Expanding Circle distinctions not as clear-cut dichotomies but as a continuum (Buschfeld & Kautzsch 2020: 56).

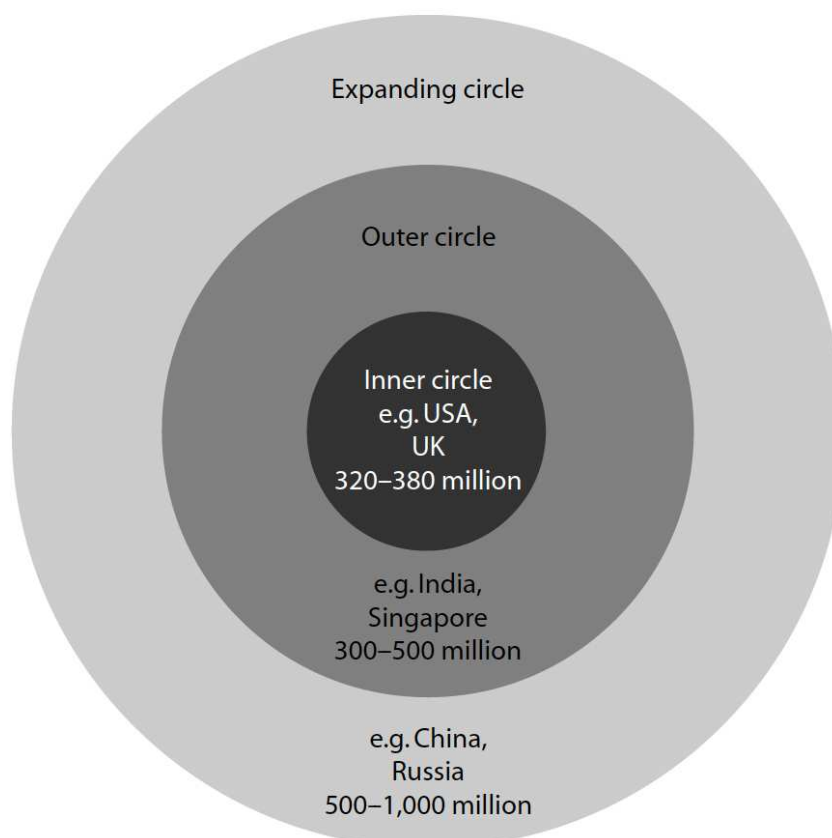


Figure 2.1: The three 'circles' of English (Crystal 2003: 61)

The same line of criticism applies to two other models which were developed at about roughly the same time, namely McArthur's (1987) "Wheel Model" and Görlach's (1988) "Circle Model of English", which both have in common that they conceptualize one variety (McArthur's "World Standard English" or Görlach's "International English") as a prestigious center hub with prescriptive power (Buschfeld & Kautzsch 2020: 56). Since these models are less prominent (for an overview, see e.g. Mesthrie & Bhatt 2008: 27-29; Buschfeld & Kautzsch 2020: 56-57), they will not be discussed further here. However, just like Kachru's approach, these models adopt a static perspective and lack a diachronic dimension (McArthur 1998: 98; Buschfeld & Kautzsch 2020: 57). Schneider's (2003; 2007) Dynamic model of the evolution of PCEs, which will be introduced in the

next section (2.2.2), was thus the first model to explicitly take into account the complex and dynamic realities that characterize the development of varieties of English.

2.2.2 *Schneider's (2003, 2007) Dynamic Model*

According to Schneider's (2003, 2007) Dynamic Model, the evolution of PCEs is tied to a specific type of contact setting involving a STL (settler) strand which brought the English language abroad in the context of British or American colonialism and an IDG (indigenous) strand which came into contact with the English language as a consequence of being colonized. Schneider therefore assumes that the emergence of PCEs can be described by a uniform evolutionary process consisting of up to five developmental stages called (1) "foundation", (2) "exonormative stabilization", (3) "nativization", (4) "endonormative stabilization", and (5) "differentiation". Each of these stages is marked by the monodirectional causal interaction of the following four parameters (see Schneider 2007: 30–31; cf. figure 2.2):

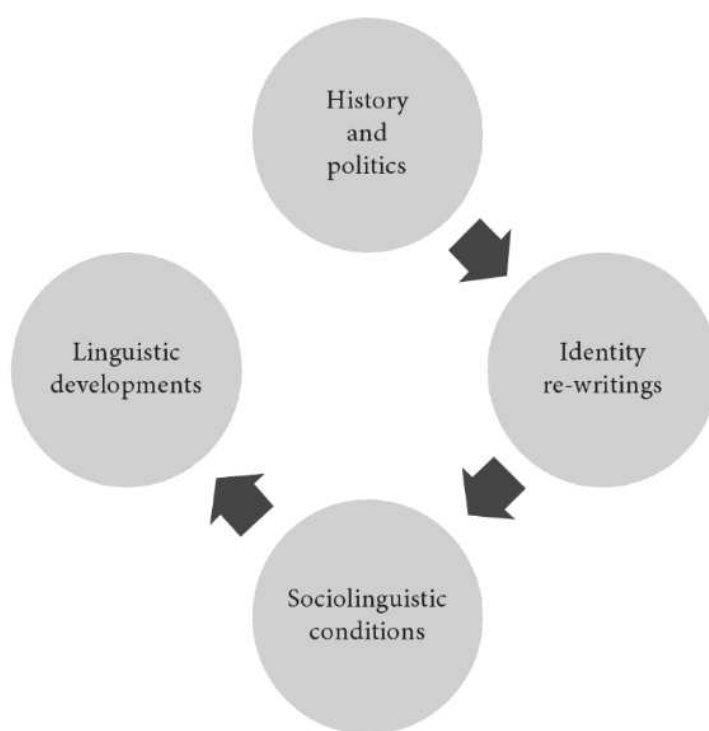


Figure 2.2: The unilateral relationship of the four sets of parameters of the Dynamic Model (Buschfeld et al. 2014: 6)

- (1) "Extralinguistic factors" such as historical and political events give rise to
- (2) "characteristic identity reconstructions" of the STL and the IDG groups, which have a direct impact on

- (3) “sociolinguistic determinants of the contact setting” such as language attitudes, language use and the intensity of the contact, which in turn lead to
- (4) “structural effects”, i.e. specific linguistic developments in the variety under discussion.

Schneider (2007: 28) thus considers “identity constructions and realignments, and their symbolic linguistic expressions” the driving force behind the emergence of PCEs. As varieties proceed along the cycle, the STL and the IDG strands increasingly become aware of a shared identity, which leads to more widespread use of English, and, ultimately, also to linguistic innovation. In the following, an overview of the five model phases will be given that briefly sketches the characteristics of each evolutionary stage. Importantly, however, not all varieties necessarily go through all five model stages. Instead, there are also countries such as the Philippines “where the in-built developmental trends of the Dynamic Model get overruled by changing external conditions, thus coming to a halt” (Schneider 2007: 143).

Phase I: Foundation

In Phase I, a group of English-speaking settlers brings English to a new territory in which English has not been spoken before. As both the STL and the IDG strands aim to maintain their own identities, English is only used for utilitarian purposes. Hence, linguistic contact and cross-linguistic influence are very limited. The most important linguistic effects in Phase I are koinéization and toponymic borrowing (Schneider 2007: 33–36).

Phase II: Exonormative stabilization

Phase II is characterized by political stabilization and the establishment of colonies, which typically involves the official introduction of English in formal domains of use such as politics, education or the legal system (Schneider 2007: 36–37). The settlers still think of themselves as belonging to their home country but gradually adopt an “English-cum-local” (Schneider 2003: 246) identity. Among the indigenous population, bilingualism starts to expand but is mostly limited to speakers with a high socioeconomic status. While the educated home variety of the settlers is clearly accepted as the general norm for spoken and written communication, the growing contact between the two groups leads to an increase in linguistic innovation (Schneider 2007: 38–40). Not only does the STL strand start to borrow terms from local languages for unfamiliar objects and concepts, but also Phase II is the beginning of “structural nativization”, i.e. “the emergence of structures which are distinctive to the newly evolving variety” (Schneider 2007: 39–40).

Phase III: Nativization

Schneider (2007: 40) describes the third phase of evolution as “the most interesting and important, the most vibrant one, the central phase of both cultural and linguistic transformation”. The driving force behind this transformation are identity reconstructions as both the settlers and the indigenous population gradually realize that they have to get along for a successful future. Frequently, this also leads to political independence. As the STL and the IDG strands start to adopt to each other, “[t]he boundary between “us” and “others” is gradually redrawn by individuals in both groups” (Schneider 2007: 41). This often goes along with discussions about old norms and the gradual acceptance of localized linguistic forms in formal contexts (Schneider 2007: 43). In addition, the growing awareness of a shared identity leads to a strong increase in the use of English, which is reflected in large-scale linguistic changes that are “at the heart of the birth of a new, formally distinct PCE” (Schneider 2007: 44). These linguistic innovations affect all levels of linguistic organization, i.e. phonology, lexis, morphology and syntax. As will be shown in 2.2.3.1, for the present study, it is particularly relevant that structural innovation typically occurs “at the interface between grammar and lexis” (Schneider 2007: 46).

Phase IV: Endonormative stabilization

In Phase IV, which normally follows political independence, the new identity constructions are strengthened. Frequently, this is triggered by “some exceptional, quasi-catastrophic political event which ultimately causes the identity alignment of STL strand speakers to switch from a self-association with the former mother country, however distant, to a truly independent identity” (Schneider 2007: 48). Schneider (2007: 49) also refers to this event as “Event X”. As the new social identity is based on collective experience and the idea of a shared territory of residence, ethnic boundaries become less important. Localized linguistic forms are gradually adopted as carriers of the new identity, and speakers generally develop a positive attitude toward the local norm, which is increasingly accepted even in formal registers. Furthermore, Stage IV sees the emergence of literary creativity and dictionaries. On the linguistic level, the strong emphasis on unity and collective identity may explain why Phase-IV varieties tend to be characterized by a high degree of homogeneity (Schneider 2007: 49–52).

Phase V: Differentiation

In Phase V, the new nation has reached stability and self-dependence, while “[t]he emergence of a new variety of English [...] is almost a thing of the past” (Schneider 2007: 53). Since the nation no longer feels the need to construct its identity as distinct from its

mother country or other entities, its citizens now define themselves as members of social or ethnic sub-groups (Schneider 2007: 53). In contrast to the previous phase, Stage V, which Schneider (2007: 54) also refers to as “the stage of dialect birth”, is thus marked by a high degree of heterogeneity as local dialects emerge as markers of social, ethnic or regional identities (Schneider 2007: 53–55).

Schneider’s model has been accepted as one of the most widely used approaches in the study of World Englishes for a number of reasons: Compared to the geography-based and static models presented in 2.2.1, Schneider’s model puts much more emphasis on the role of social aspects in linguistic developments (Van Rooy 2010: 10). Moreover, it incorporates a diachronic dimension as it emphasizes that social, political, and linguistic realities are constantly changing (Edwards 2016: 6). In addition, Schneider makes predictions about the linguistic effects that are associated with different evolutionary stages, which can be assessed using corpus-based research (Brunner 2017: 18).

Nevertheless, there are also several points of criticism: It has been argued that the model puts too much emphasis on the occurrence of a political “Event X”, which may be of little relevance for the identity of individual speakers, particularly with respect to lower classes. The model thus conceptualizes identity mostly in terms of nationhood, thereby ignoring aspects like class or status, which can also be expected to contribute to identity constructions (Mesthrie & Bhatt 2008: 35). Mair (2017) criticized that the model does not take into account transnational developments. Furthermore, the model has been accused of overgeneralizing: The fact that Schneider includes “differentiation”, i.e. the emergence of local dialects as markers of group identities, only in the final stage of his model fails to account for varieties that have not yet developed a clear endonormative orientation but still subsume various subdialects that serve as identity markers (Mesthrie & Bhatt 2008: 35). For example, with respect to NigE, which is categorized as a Phase-III variety by Schneider (2007: 210), Gut (2004: 815) pointed out that “the diversity of the different kinds of English in the country is so great that Nigerian English (NigE) is usually divided into several sub-varieties” that are associated with different ethnic groups. It is important to note, though, that Schneider (2007: 56) himself underlined that the Dynamic Model is a theoretical model which aims to point out similarities between independent developments that are helpful for making predictions. As this necessarily requires generalizations and abstractions that do not cover all details and complexities of real processes, Schneider (2007: 55) emphasized that a theoretical model such as the Dynamic model “should not be confused with reality itself”.

For the present study, which aims to gain insights into the cognitive aspects that characterize the evolution of PCEs, the Dynamic Model assumes great importance. As the next section will show, Schneider's sociolinguistic account of the emergence of New Englishes can also be explained from a cognitive perspective.

2.2.3 *Hoffmann's (2021) Construction Grammar Approach to the Dynamic Model*

Drawing on basic assumptions of Usage-based Construction Grammar, Hoffmann (2014, 2019, 2021) added a cognitive dimension to Schneider's (2003, 2007) model. In his "Construction Grammar Approach to the Dynamic Model", Hoffmann (2021) does not just provide a constructionist explanation of the linguistic effects associated with the five model stages, but he also predicts that the evolutionary stage of a variety should be correlated with the productivity of constructional slots. Laporte (2021: 75) praised Hoffmann's approach for "complement[ing] the social dimension of the Dynamic Model where structural nativization is seen mostly as a result of identity-construction and norm-orientation with one that also considers the linguistic representation of speakers". In a similar vein, Hoffmann (2021: 14) himself pointed out that

a reinterpretation of the linguistic innovations of PCEs as constructional innovation is not only merely a case of re-labelling existing analyses. Instead, the constructionist approach provides a cognitively plausible account of the linguistic evolution of PCEs and, on top of that, also raises new questions for future studies within the Dynamic Model.

As Hoffmann's constructionist interpretation of the Dynamic Model is highly relevant for the corpus study in chapter 5, in this section, the main aspects of his approach will be dealt with in more detail.

2.2.3.1 Qualitative innovations from a constructionist perspective

According to Hoffmann (2021: 14), the most important linguistic changes in early phases of Schneider's model tend to be situated at a low level of abstraction. From a Construction Grammar point of view, toponymic borrowing, which is characteristic of Phase I, means that fully substantive constructions such as place names, which are usually straightforward form-meaning pairings with a high token frequency, become entrenched in the construction of the settlers (Hoffmann 2021: 14). The entrenchment of substantive constructions also plays a major role in Phase II, in which "a set of fully substantive, locally salient and very frequent micro-constructions" (Hoffmann 2021: 17) is added to the construction network of the STL strand. Compared to the place names that become entrenched in the Foundation stage, these form-meaning pairings typically refer to objects (Hoffmann 2021: 17).

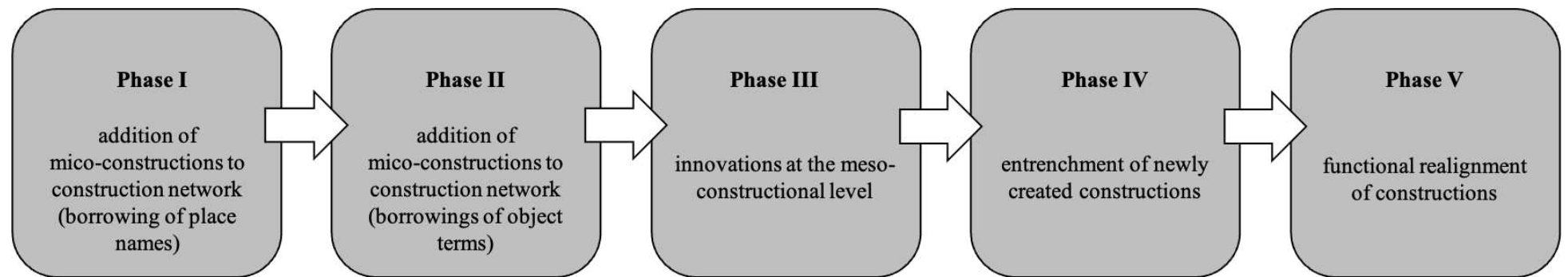


Figure 2.3: Linguistic innovations in the Dynamic Model from a constructionist perspective (based on Hoffmann 2021)

In contrast, in Phase III – the most significant evolutionary stage in terms of linguistic innovation (cf. 2.2.2) – constructional change is not restricted to the addition of micro-constructions anymore. According to Schneider (2007: 83), innovations that emerge during Nativization

are frequently positioned **at the interface between lexis and grammar**, i.e. certain words but not others of the same word class prefer certain grammatical rules or patterns. The patterns as such are not new, nor are the words, but what is novel is the habitual association between them in specific varieties. (Schneider 2007: 83; own emphasis)

Hoffmann (2021: 19) thus claims that structural innovation in PCEs typically affects “the meso-constructional level”, i.e. partly schematic constructions that contain elements that are phonologically fixed as well as schematic slots that can be filled by various expressions. Examples of such idiosyncratic innovations at the meso-constructional level include changes in verb complementation patterns (e.g. the ditransitive use of *advise* or *impart* in IndE; Mukherjee & Hoffmann 2006: 164) or variation in the use of prepositions (Schneider 2007: 46–47). Jowitt (2019: 92), for example, observed that NigE has a preference for complex prepositions such as *since on* or *until on* (as in *Do not pay him until on Wednesday*) in cases in which the preposition slot would simply be filled by *since* or *until* in Standard Englishes. The emergence of novel verb-preposition combinations (e.g. *indulge into* instead of *indulge in*), which is characteristic of PCEs (cf. 2.4.2), can also be said to start at the meso-constructional level as it involves constructions that consist of a verb slot that is fully specified (e.g. *indulge*) as well as a preposition slot that is not fixed (for examples of innovations affecting the preposition slot of verb/adjective-preposition strings, also see e.g. Huber & Dako 2004: 861 on Ghanaian English; Budohoska 2014: 126; Buregeya 2019: 93 on KenE; Balasubramanian 2009 on IndE; Deterding 2007: 57 on SgE; Hundt, Hay & Gordon 2004: 579 on NZE; Schneider & Gilquin 2018). Furthermore, Laporte’s (2021: 261) study on HKE, IndE and SgE found “subtle” variation in the preposition slot of partly schematic constructions involving the verb *make*, which leads to novel patterns such as *WHAT X MAKE OUT OF Z* (e.g. *I don’t know what to make out of that*) instead of the standard variant *WHAT X MAKE OF Z* (e.g. *I don’t know what to make of that*). Consequently, there is ample evidence which shows that linguistic innovation in Stage III typically affects partly schematic constructions.

According to Hoffmann (2021: 19), the assumption that idiosyncratic innovations in PCEs surface at the meso-constructional level is consistent with research on first language acquisition, which found that children’s first instances of creative language use typically also involve partly schematic constructions (see Diessel 2013: 352). Braine & Bowerman (1976), for example, showed in a study of English, Samoan, Finnish, Hebrew

and Swedish child language that the first multiword constructions in L1 acquisition (e.g. *want milk*, *want car* or *more music*, *more cookie*; Braine & Bowerman 1976: 44) normally consist of a “pivot word”, such as *want* or *more*, which is fixed, as well as an “open slot”, which can be filled freely (Diessel 2013: 352; also see Tomasello 1992). Likewise, children’s early yes-no questions such as *Can I X?* typically include a fixed auxiliary, a pronoun slot, and a verb slot that can be filled by various lexical fillers (Dąbrowska 2000: 91–92). Note that these parallels between the creative language use of children and the emergence of structural nativization in PCEs also provide support for Schneider’s (2007: 110) general claim that the evolution of PCEs is “shaped by the same conditions of language change and evolution as any other language”. Importantly, however, this does not imply that speakers of PCEs only store meso-constructions. Although linguistic innovation typically surfaces in partly schematic constructions, Hoffmann (2021: 19) suggested that L2 speakers might even be able to make faster generalizations to more abstract constructions than L1 speakers since their cognitive skills should be more advanced than those of children.

While the third phase of the Dynamic Model is thus characterized by complex changes at an intermediate level of the construction network, in Phase IV, which is generally marked by a high degree of homogeneity, the newly created constructions are entrenched and the construction network is stabilized. The final model stage, in which dialects emerge as markers of group identities, then sees “the functional realignment and innovation of constructions as linguistic means of identity” (Hoffmann 2021: 20). Hoffmann’s (2021) constructionist view of linguistic innovations in the Dynamic Model is also summarized in figure 2.3.

2.2.3.2 Hoffmann’s (2021) Dynamic Model Productivity hypothesis

In addition to the qualitative effects described in the previous section, Hoffmann (2021: 21) also predicts quantitative differences between variety phases that can be expected to arise due to the specific type and amount of input that is associated with different developmental stages. In earlier Dynamic Model phases, in which the use of English is restricted to fewer domains (cf. 2.2.2), speakers can be expected to encounter less input for constructions. Since the entrenchment of abstract constructions depends on input frequency (cf. 1.1.1), the cognitive constructional representations of speakers of less advanced varieties should thus not be as abstract as those of speakers of varieties at higher developmental stages. Hoffmann (2021: 21) summarized this correlation between the

evolutionary stage of a variety and the productivity of constructional slots in his “Dynamic Model Productivity hypothesis”:

Dynamic Model Productivity hypothesis: Varieties in later phases of the Dynamic Model show (1) more productivity of the slots of (semi-)schematic construction than varieties in earlier developmental phases due to (2) less reliance on prototypical and frequent fillers.

Support for the claim that the construction networks of Phase-V varieties contain more abstract meso- and macro-constructions than those of varieties at earlier Dynamic Model stages comes from a number of recent studies such as Hoffmann’s (2014, 2019) exploration of the Comparative Correlative construction. The Comparative Correlative construction (e.g. *the more you read, the more you know*) consists of two clauses (*the more you read – the more you know*) that both contain the fixed element *the* as well as schematic slots (Hoffmann 2019: 1–2). Analyzing the Comparative Correlative construction in 12 ICE components, Hoffmann (2014: 175, 2019: 224) found that Phase-III varieties use more specific lexical filler pairs such as *more...more* than more advanced varieties. While three such partly schematic constructions emerged as being significantly associated with PhilE, only one such partly schematic construction turned out to be entrenched in the Stage-IV variety JamE, and not a single meso-construction was identified as being significantly associated with BrE. Hoffmann (2019: 225) thus concluded that “stage V PCEs make greater use of a macro-construction”. This effect was corroborated by Horsch’s (2023: 184) investigation of the Comparative Correlative construction in the Corpus of Global Web-based English (GloWbE), which found that entrenched meso-constructions make up a larger proportion of the comparative correlative tokens in Phases III and IV than in Phase V.

In addition, there is a series of studies using so-called “Large Number of Rare Events” (LNRE) modeling (Baroni & Evert 2014; Zeldes 2013), which can be applied to predict the productivity of constructional slots by extrapolating productivity curves beyond the observed sample size (for more details about the advantages of this technique, see section 4.4.3). Drawing on GloWbE data, Brunner & Hoffmann (2020) used this statistical method to investigate the productivity of the verb slot of the *way*-construction (e.g. *She **worked** her way to the top*) across Dynamic Model stages. In line with Hoffmann’s (2021) prediction, the slot productivity turned out to be correlated with evolutionary stages, with Phase V exhibiting the highest and Phase II exhibiting the lowest productivity (Brunner & Hoffmann 2020: 15). Furthermore, Brunner & Hoffmann (2020: 16) conducted an HCFA (see section 4.4.2) to test whether certain lexical fillers are significantly associated with certain Dynamic Model phases. In accordance with the second part of the Dynamic Model Productivity hypothesis, they were able to show that

varieties at earlier stages strongly favor the prototypical filler *find* (as in *She **found** her way to the door*; Brunner & Hoffmann 2020: 16). Moreover, with respect to the preposition slot, Phase III turned out to have a stronger preference for *to* than Phase V. The meaning of *to* being highly compatible with the prototypical meaning of the *way*-construction (i.e. completed motion), this implies that less advanced varieties tend to rely more on prototypical slot fillers than varieties at higher stages (Brunner & Hoffmann 2020: 19–20; Hoffmann 2021: 33–34).

Another study which found a clear correlation between Dynamic Model phases and the productivity of constructional slots is Hoffmann’s (2020a) LNRE analysis of the verb slot in the [*V the N_{taboo-word} out of*]-construction (e.g. *You scared the hell out of me*). Just like in the analysis of the *way*-construction, the V slot turned out to be least productive in Phase II and most productive in Phase V, while Phases III and IV occupied the middle ground (Hoffmann 2020a: 6). In addition, Hoffmann (2021: 37–39) conducted an HCFA of the noun slot, which identified three significant taboo nouns (e.g. *hell*) that are significantly associated with Phase V, but only two for Phase IV, and one for Phases II and III. As Hoffmann (2021: 37) pointed out, these quantitative differences can be taken as an indication that “the taboo N constructions generally increase in use across the Dynamic Model phases”, which is compatible with the assumption that speakers of more advanced varieties receive more input for the construction.

A third construction that was studied by Hoffmann (2021: 44–46) with the help of LNRE modeling is the [*as ADJ as a N*]-construction (e.g. *as fit as a fiddle*). With respect to the productivity of the combined ADJ-N slot (e.g. *fit-fiddle*), the results again support the view that more advanced varieties have more abstract constructional representations as the study found a clear cline from Stage V, which emerged as the most productive variety type, to Stage II, which exhibited the lowest slot productivity (Hoffmann 2021: 45).

Support for the second part of Hoffmann’s (2021) hypothesis also comes from Bruckmaier’s (2017) corpus-based study of GET-complementation in ICE-Great Britain, ICE-Jamaica and ICE-Singapore, which suggested that SgE and JamE have a stronger preference for the prepositional object construction (GET *something for somebody*) than BrE, while they tend to avoid the less iconic ditransitive use of GET (GET *somebody something*). Bruckmaier (2017: 84) interpreted this finding as “a clinging to [...] simple and familiar patterns” in L2 varieties which can be attributed to second-language acquisition effects (cf. 2.4.2). Edwards & Laporte (2015: 151–152) analyzed the pattern V + *into* (e.g. *come/turn/break/look into*) in ENL, ESL and EFL varieties, finding that

HKE and Dutch English exhibit less variation in the verb slot than IndE and SgE, i.e. varieties they considered more institutionalized than HKE. In addition, they observed that speakers of HKE and Dutch English tend to rely more on (semi-)fixed expressions, which can be interpreted as a “tendency of learners to stick to expressions they know and consider safe” (Edwards & Laporte 2015: 155). Furthermore, Laporte’s (2021: 321) study on the verb *make* in BrE, HKE, IndE and SgE indicated that more advanced varieties show more variation in the use of abstract constructions than varieties at lower stages of evolution. At the same time, HKE, for example, relies more on a small set of partly schematic light verb constructions such as *make a decision* than more advanced varieties (Laporte 2021: 318).

Importantly, however, there are also studies that do not support the Dynamic Model Productivity hypothesis. Horsch’s (2023: 187–190) LNRE models assessing the productivity of constructional slots (e.g. *more-more*) in the Comparative Correlative construction indicated that there is “little difference in productivity between [Dynamic Model] stages” (Horsch 2023: 187). Likewise, Brunner’s (2021) investigation of the *into*-causative construction (e.g. *She talked him into buying a car*) in 20 GloWbE varieties did not find significant differences between variety phases with respect to the productivity of the cause slot (e.g. *talk*) and the result slot (e.g. *buying*) as well as the filler types used in these slots, which suggests that all varieties have entrenched similar meso- and macro-constructions (Brunner 2021: 16–25). Brunner (2021: 26) speculated that the differing results between his study and e.g. Brunner & Hoffmann’s (2020) analysis of the *way*-construction might be due to the fact that the interpretation of *into*-causatives, whose meaning is relatively straightforward, is not as complex as that of the *way*-construction, in which the verb, the preposition and *way* have to be interpreted in a non-literal sense. Moreover, he pointed out that GloWbE, on which his study was based, does not include truly informal registers, in which variety-specific effects should be most likely to surface (Brunner 2021: 26–27). The conflicting results of the studies conducted so far thus indicate that more construction types have to be investigated in order to assess the validity of Hoffmann’s predictions.

Finally, it is important to note that the Dynamic Model Productivity hypothesis does not mean to suggest that the construction networks of all varieties of the same phase contain exactly the same meso- and macro-constructions, which would imply that all varieties that have reached Phase V end up sharing the same constructicon (Hoffmann 2021: 22). Instead, Hoffmann’s (2021) hypothesis only makes predictions about a correlation between Dynamic Model stages and quantitative linguistic effects at a high

level of abstraction, namely the productivity of constructional slots. The Dynamic Model productivity hypothesis is thus fully compatible with the assumption that the construction networks of PCEs exhibit qualitative linguistic innovations such as novel verb-preposition patterns (Hoffmann 2021: 46; cf. 2.2.3.1).

2.3 Stylistic variation in PCEs from a cognitive perspective

Another aspect that characterizes the evolution of PCEs is variation at the stylistic level. There is a large body of research suggesting that L2 Englishes tend to use “many typically written features in their speech and typically spoken features in their writing” (Deshors, Götz & Laporte 2018: 14). This effect has also been referred to as a “lack of register awareness” (Gilquin & Paquot 2008) in New Englishes. However, as this term implies that PCEs fail to reach a desired standard of awareness and are thus somehow inferior to Standard varieties, in the present study, the more neutral term “lack of register distinctions” (Mußemann 2024: 2) will be preferred. This section aims to provide a cognitive explanation of this effect.

From a Construction Grammar point of view, stylistic innovation can be considered the result of “a restructuring in form-function mapping” (Deshors, Götz & Laporte 2018: 13). Stylistic innovation in PCEs can be found at all levels of abstraction. For instance, Buregeya (2019: 171) found that KenE speakers tend to use formal terms such as *alight* instead of more informal variants such as *get off*, which suggests that KenE has a preference for micro-constructions that are associated with formal contexts in BrE. Likewise, there are meso-constructions which are restricted to formal registers in BrE but which are also used in informal contexts in New Englishes. For example, Bernaisch (2015: 87) argued that a BrE speaker would interpret the positioning of *earlier* right before the verb in example (2.1) as a signal of formality, while the adverb would be placed clause-initially in a stylistically unmarked construction in BrE. In contrast, speakers of IndE and Sri Lankan English use *early* in pre-verbal position even in informal contexts. As the construct in (2.1) is licensed by a meso-construction which contains the fixed element *earlier* as well as open slots that can be filled by various lexical fillers, this is an example of a restructuring in the form-function mapping of a meso-construction (example in (2.1) from Bernaisch 2015: 87; original emphasis). An example that demonstrates that stylistic innovation can also affect macro-constructions is Hoffmann’s (2011: 267) finding that KenE strongly favors the pied-piped finite WH-relative clause construction across all registers, while the construction is associated with formal contexts in BrE. As

this effect is highly relevant for the present study, it will be discussed again in section 3.3.1.

- (2.1) *This place **earlier** was for her alone – but today Duleep is sitting in front of her, and she is trying to adjust her mind to presence in her space.* <ICE-SL:W2F-008#22:1>

While the majority of the studies that found a lack of register distinctions in PCEs have focused on individual linguistic phenomena or varieties (see e.g. Shields-Brodber 1989: 46 on JamE; Gonzalez 1991 on PhilE; Balasubramanian 2009: 228 on IndE; Van Rooy et al. 2010: 346 on East African English; Jowitt 2019: 144–145 on NigE), Bohmann (2019: 194) used a feature-aggregational method to study various dimensions of register variation in the ICE corpora, concluding that (also see Xiao 2009: 441–446)

[t]he general picture that crystallizes from a joint consideration of several factors is that countries in earlier stages of post-colonial linguistic identity-formation tend to be more formal and informational, and less affective and involved than their phase 5/Inner Circle counterparts. This is evidenced particularly in the spoken corpus data by higher scores of phase 3 and 4 countries on the dimensions of INVOLVED VERSUS INFORMATIONAL PRODUCTION, CONCEPTUAL VERSUS CONCRETE INFORMATIONAL FOCUS, and ASSERTION OF FACTUAL VALIDITY as well as lower scores on dimension EXPLICIT STANCE-MARKING.

At the same time, the literature seems to suggest that there is a correlation between the degree of stylistic variation and the evolutionary stage of a variety. Bao and Hong (2006: 112) found clear evidence of stylistic variation in SgE, which is classified as a Stage-IV variety by Schneider (2007: 153). Likewise, Brunner (2017: 287) concluded that “speakers of Singaporean English [...] largely shed their preference for adopting simpler NPs than those in the native variety when writing formal scholarly texts [...], [while] the tendency remains in place in Kenyan English”, which is usually considered a Stage-III variety (Schneider 2007: 192–197). In a study on relative clauses, Gut & Coronel (2012: 238) also found more signs of stylistic variation in more advanced varieties, arguing that “in the two New Englishes that show the clearest exonormative orientation [, i.e. NigE and PhilE], stylistic variability is lower than in the other two [, i.e. JamE and SgE]”. In a similar vein, Hundt (2006: 219) claimed that “PhilE is subject to a strong exonormative pressure that even carries over to spoken usage at times, whereas SingE can be seen as developing the same kind of stylistic variation found in inner-circle varieties”. These examples show that norm orientation is frequently considered the main reason behind differences in the degree of stylistic variation between more and less advanced varieties. However, as norm orientation may not necessarily have a strong impact on the linguistic choices of individual speakers (also see Laporte 2021: 332–333 for the role of norm-orientation in the evolution of PCEs), I will put forward a cognitive explanation that

shows that the apparent correlation between the strength of stylistic effects and the evolutionary stage of a variety can be attributed to input differences between lower and higher Dynamic Model stages.

In order to explain this effect, it is helpful to briefly introduce Schmid's (2020) Entrenchment-and-Conventionalization Model, which understands language as "a structured and yet highly dynamic and constantly changing system" (Schmid 2020: 3) that is "driven by speakers' linguistic activities in repeated usage events" (Schmid 2020: 4). At the heart of Schmid's model lie the processes of "entrenchment" and "conventionalization". While entrenchment describes the "routinization" and "schematization" of linguistic knowledge in the minds of individual speakers, conventionalization refers to the "usualization" and "diffusion" of utterances at the community level (Schmid 2020: 5–6). Both processes, which constantly interact and influence each other, are strongly driven by specific usage situations. With respect to register distinctions, this means that, as a result of specific usage events, pragmatic associations that link constructions to specific contexts become routinized and entrenched in the minds of individual speakers (Schmid 2020: 208–209). At the same time, constructions can become conventionalized as the appropriate choice in certain contexts at the community level through repeated usage events in which these pragmatic associations are activated (Schmid 2020: 5–6). An important mechanism in the learning of such pragmatic associations is a process known as "preemption" (Tomasello 2003: 300; Goldberg 2006: 94–98). The central assumption behind preemption is that "if someone communicates to [someone] using Form X, rather than Form Y, there was a reason for that choice related to the speaker's specific communicative intention" (Tomasello 2003: 300). Preemption thus describes a speaker's effort to make generalizations about the functional distinction of constructions based on the input the speaker is exposed to.¹

But how do these processes and the learning of pragmatic associations play out with respect to stylistic variation in New Englishes? First of all, note that speakers of varieties at earlier stages of Schneider's cycle mostly receive input for formal constructions since PCEs are typically used in formal domains, while other languages are preferred for more private matters (Van Rooy et al. 2010: 346). Furthermore, PCEs are

¹ Note that the concept of preemption is frequently related to Goldberg's (1995: 67) "Principle of No Synonymy", which claims that constructions that are syntactically distinct must be distinguished by pragmatic or semantic differences. However, e.g. Uhrig (2015) showed that preemption is not necessarily compatible with the Principle of No Synonymy.

usually acquired through formal education, where formal and even archaic constructions tend to be very common. As Van Rooy (2010: 14) pointed out, apart from non-native teacher language, particularly written teaching material, which is typically marked by a very formal style, can be considered one of the main sources of input in New Englishes. The frequent observation of a tendency towards more formal variants in Stage-III varieties, such as KenE, PhilE or NigE, can thus be explained by the fact that these speakers mostly receive input for constructions that are associated with formal registers in BrE (or, e.g. in the case of PhilE, American English). In contrast, in the more advanced Stage-IV varieties, such as SgE or JamE, English is used in a wider range of domains, which means that these speakers are not only exposed to more input but also to a higher type frequency of constructions. In accordance with Hoffmann's (2021) predictions (cf. 2.2.3.2), it can therefore be assumed that speakers of Stage-IV varieties have more entrenched meso- and macro-constructions than speakers of less advanced varieties. While the construction networks of Stage-III varieties only contain a limited set of meso- and macro-constructions, namely the formal constructions that surface in the formal input these speakers are exposed to, more advanced varieties possess more entrenched constructional alternatives from which they can choose.

Now, considering that speakers always seek to identify functional differences between constructions, it can be assumed that the constructions that are entrenched in the mental grammars of Phase-IV speakers are associated with specific contexts. Just like the Phase-III varieties, the Phase-IV varieties were also frequently exposed to constructions that are considered formal in BrE in the formal domains of use through which English was introduced in the British colonies. In accordance with Schmid's (2020) Entrenchment-and-Conventionalization Model, this most likely led to the entrenchment of pragmatic associations between these formal constructions and formal usage contexts. Due to the influence of statistical preemption, alternative constructions that became more deeply entrenched over the course of time as the use of English expanded, then became associated with more informal situations, which means that pragmatic associations that link these constructions to informal contexts became routinized and entrenched in the minds of individual speakers (Mußemann 2024: 24). As a result of these pragmatic associations, certain constructions are much more likely to be activated in certain situations than constructional alternatives that are associated with other registers (Schmid 2020: 6). At the same time, repeated usage events contributed to the conventionalization of specific constructions in specific contexts at the community-level.

In contrast, speakers of Phase-III varieties possess fewer alternative constructions from which they can choose. The limited degree of stylistic variation in Phase III can thus be explained by the fact that speakers of varieties at lower stages, who receive less input, typically have fewer entrenched constructions than speakers of more advanced varieties. Support for this claim comes from research on L1 acquisition, which suggests that statistical preemption can only begin to play a role after children have acquired a set of alternative constructions. For instance, with respect of the acquisition of novel verbs, Tomasello (2003: 180) pointed out that,

[entrenchment] is the only factor at work for some time, as children do not have many alternative constructions to effect preemption, and they have not yet had time to learn semantic subclasses of verbs. Preemption and semantic subclasses begin to work later, perhaps not until 4;6 or so, as children learn more about the conventional uses of verbs and about all of the alternative linguistic constructions at their disposal in different communicative circumstances. Thus, just as utterance-level constructions become more abstract only gradually and in piecemeal fashion, so also are they constrained only gradually and in piecemeal fashion.

In L1 acquisition, preemption therefore only works when children possess constructional alternatives. In a similar vein, it is plausible to conclude that L2 speakers only begin to associate specific constructions with specific contexts after they have received sufficient input that allows them to acquire alternative constructions. From a cognitive perspective, the lower degree of stylistic variation in earlier Dynamic Model phases can thus be attributed to the amount and type of input that is typical of these varieties. With respect to the present study, it can therefore be predicted that varieties at higher stages of evolution should exhibit a greater degree of stylistic variation than varieties at lower developmental stages.

Finally, note that, even though speakers of less advanced varieties primarily encounter formal input, in some cases, the lack of alternative constructions may also lead to an overuse of informal variants in formal contexts (see e.g. Gilquin & Paquot 2008). For example, Brunner's (2017: 288) finding that KenE speakers tend to use simple noun phrases (NPs) even in formal academic texts, which are characterized by more complex NP patterns in BrE, can be explained by the fact that the use of complex NP constructions, which requires more processing effort (Brunner 2017: 289–290), is less deeply entrenched in KenE. Consequently, these speakers also rely on simple NP constructions in formal contexts.

2.4 Other relevant factors in the evolution of PCEs

While the previous sections provided a cognitive explanation of the linguistic changes (cf. 2.2.3) and stylistic effects (cf. 2.3) that shape the evolution of PCEs, this section will present other factors that are assumed to influence the emergence of New Englishes.

2.4.1 *Processing factors*

One aspect that has been frequently mentioned in the literature is the role of processing factors. As pointed out in 1.1, particularly interesting for the present analysis is Hoffmann's (2011) study on preposition stranding and pied-piping in BrE and KenE, which suggested that KenE speakers have a stronger preference for preposition placement constructions that involve a reduction in processing cost than BrE speakers. Drawing on basic assumptions of Usage-based Construction Grammar, Hoffmann (2011: 270) linked this stronger preference for prototypical constructions in the L2 variety to input. Remember that usage-based approaches emphasize that abstract constructional representations can only emerge if speakers encounter a construction with a high type frequency, while substantive constructions can become entrenched as a result of a high token frequency. As a consequence, it can be assumed that the construction networks of L2 speakers, who are exposed to less input, are not as deeply entrenched as those of L1 speakers, which means that constructions cannot be processed as fast and as easily (cf. 1.2.1). In L2 varieties, the higher cognitive cost associated with the use of English may thus lead to a stronger preference for constructions that are easy to process (Mußemann 2024: 5). Hoffmann (2011: 275) describes this correlation between the degree of constructional entrenchment and the strength of processing constraints as follows:

[M]ost Kenyan English speakers will possess a construction network that is less deeply entrenched than that of L1 British English speakers [...]. As a result, Kenyan English speakers are to a greater degree subject to processing constraints during online production. Consequently, it is not surprising that such structures [that involve a processing advantage] appear more frequently in the second-language learner's data [...] and receive better scores from Kenyan English speakers [...].

While the differences between the KenE and the BrE construction networks which Hoffmann (2011) attributed to processing factors will be outlined in 3.3.1, the first part of this section (2.4.1.1) will discuss the results of other corpus studies that explored the effect of processing constraints on speakers' linguistic choices in Englishes around the world. It will be shown that the findings generally indicate that processing-related differences between L1 and L2 speakers are most likely to emerge in studies focusing on alternating constructions whose competition is strongly guided by parsing-related variables and which do not exhibit major semantic differences. The second part of this

section (2.4.1.2) will give an overview of psycholinguistic research that provides support for the claim that processing constraints have a stronger effect on L2 speakers than on L1 speakers.

2.4.1.1 Processing factors in World Englishes

That L2 speakers are, at least in parts, to a greater degree subject to processing factors than L1 speakers was not only suggested by Hoffmann's (2011) investigation of preposition placement but also by studies on *that*-complementation and particle placement. Analyzing *that*-complementation (*I thought you would win* vs. *I thought **that** you would win*) in the German and Spanish components of the International Corpus of Learner English, Wulff, Gries & Lester (2018: 117–118) found that German and Spanish L2 speakers only omit the *that*-complementizer in contexts that are easy to process, and that they generally show a higher preference for the more explicit variant. While this effect may, at least to a certain degree, be favored by cross-linguistic influence (as both German and Spanish are less permissive with respect to complementizer optionality; see Wulff 2016; Wulff, Gries & Lester 2018: 101–103), it can also be taken as an indication that L2 speakers are more strongly affected by processing constraints than L1 speakers (also see Wulff, Lester & Martinez-Garcia 2014: 291). Assuming that multilingualism is associated with higher cognitive cost (see e.g. Kroll & Dussias 2013), Wulff, Gries & Lester (2018: 118–119; also see Wulff, Lester & Martinez-Garcia 2014: 293) argued that L2 speakers dispose of fewer cognitive resources than L1 speakers. As a result, they seem to use the *that*-complementizer, which is never ungrammatical and thus a “safe” (Wulff, Gries & Lester 2018: 118) strategy, more frequently than the BrE L1 speakers. From a processing perspective, the use of the complementizer also complies with Hawkins (2004: 51) principle of “Maximize On-line Processing”, which can be summarized by the slogan “Express it earliest” (Hawkins 2004: 49). According to Hawkins (2004: 51),

[t]he human processor prefers to maximize the set of properties that are assignable to each item X as X is processed [...]. The maximization difference between competing orders and structures will be a function of the number of properties that are unassigned or misassigned to X in a structure/sequence S, compared with the number in an alternative.

In example (2.2a), in which *that* is omitted, it is likely that *the boy*, which is semantically compatible with the matrix clause *I believe*, is mistakenly interpreted as the direct object of *believe*, which leads to a garden path once *knows* is parsed. In contrast, in (2.2b) such a misassignment is avoided (examples from Hawkins 2004: 52).

- (2.2) (a) I_{vp}[*believe* [*the boy knows the answer*]]
 (b) I_{vp}[*believe* [*that the boy knows the answer*]]

Since the use of *that* thus offers a processing advantage as it avoids garden path effects and delays in the recognition of the *that*-clause (Hawkins 2004: 51–53), it is not surprising that learners of English show a stronger preference for the use of the complementizer than BrE speakers.

Moreover, several studies on particle placement found that L2 speakers tend to overuse the verb-particle-object (VPO) construction (e.g. *John **picked up** the book*) in contexts in which L1 speakers would choose the verb-object-particle (VOP) construction (e.g. *John **picked** the book **up***; Gilquin 2015: 64–65 on French-speaking L2 learners; Grafmiller & Szmrecsanyi 2018: 405 on nine ICE varieties; Paquot, Grafmiller & Szmrecsanyi 2019: 86 on seven learner Englishes; Wulff & Gries 2019: 900–901 on L2 learners from 17 language backgrounds). This effect can also be linked to a processing principle. In the VPO construction, in which the particle appears directly after the verb, the speaker can immediately establish a relation between the two elements (e.g. between *pick* and *up*). In contrast, in the VOP variant, the verb has to be held active in memory while the direct object is being processed (Gries 2003: 56; Hawkins 2004: 27; Wulff & Gries 2019: 901). The fact that the VPO variant is strongly preferred in L2 Englishes thus indicates that these speakers favor constructions that involve a reduction in processing cost. Note, however, that this does not mean that processing principles do not also guide the constructional choices of L1 speakers. Instead, several studies showed that speakers of all varieties tend to disfavor the split VOP construction with long direct objects, which require the parsing of additional material before the verb and the particle can be related (Szmrecsanyi et al. 2016: 120; Grafmiller & Szmrecsanyi 2018: 403; Paquot, Grafmiller & Szmrecsanyi 2019: 87; Wulff & Gries 2019: 902). This effect is in accordance with Hawkins' (2004: 31) "Minimize Domains" principle, which predicts that "[t]he human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed". At the same time, the stronger preference for the VPO construction in L2 Englishes mostly becomes visible in environments in which the processing load is reduced, i.e. in constructions with short direct objects (Szmrecsanyi et al. 2016: 122; Paquot, Grafmiller & Szmrecsanyi 2019: 81–82). Also note that, for example, Wulff & Gries (2019: 900) emphasized that their findings generally "paint a complex picture in which processing demands, input effects, and L1 typology jointly shape[...] the degree to which learners' choices of constructions [a]re nativelike or not" (also see Wulff, Gries & Lester 2018: 117–118; Paquot, Grafmiller & Szmrecsanyi 2019). This shows that more research is necessary to shed more light on the

factors that influence speakers' choices between competing constructions in different varieties.

While studies on *that*-complementation and particle placement provide at least some support for Hoffmann's (2011) claim that L2 Englishes have a stronger preference for prototypical constructions than L1 Englishes, research on the genitive (Gries & Wulff 2013; Szmrecsanyi et al. 2016; Heller, Szmrecsanyi & Grafmiller 2017) and the dative alternation (Bernaisch, Gries & Mukherjee 2014; Röthlisberger, Grafmiller & Szmrecsanyi 2017) concluded that processing factors affect L1 and L2 speakers in a very similar way. With respect to the genitive alternation, i.e. the competition between the N-Possessive construction (e.g. *John's book*) and the PP_{OF}-Possessive construction (e.g. *the book of John*; Hoffmann 2022: 79), Gries & Wulff (2013) explored written data from L1 British speakers, L2 Chinese learners and L2 German learners of English, finding that differences in the constructional choices between these speaker groups primarily result from cross-linguistic influence (Gries & Wulff 2013: 349–351). At the same time, they demonstrated that both L1 and L2 speakers exhibit a preference for the *of*-genitive with long possessor NPs (e.g. *the tail of the dog that was afraid of sheep* is preferred over *the dog that was afraid of sheep's tail*) and a preference for the *s*-genitive with specific possessors (e.g. *John's goals* vs. *the goals of John*). While the former effect can be explained by the principle of “end-weight”, which assumes that speakers prefer to use short constituents before long ones (Behaghel 1909/1910; cit. in Gries & Wulff 2013), the latter finding corresponds to the principle “given-before-new” (Gundel 1988), which assumes that the early use of given information can provide a reference point for the hearer that facilitates the interpretation of new material (Gries & Wulff 2013: 348). Gries & Wulff (2013: 349) therefore came to the conclusion that “intermediate/advanced learners' production data already reflect target-like processing preferences at an astonishing level of complexity”. Dubois, Paquot & Szmrecsanyi (2022) did not only corroborate these findings for spoken learner language, but they also 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 (Dubois, Paquot & Szmrecsanyi 2022: 20).

While the studies mentioned so far focused specifically on learner Englishes, Heller, Szmrecsanyi & Grafmiller (2017) explored the genitive alternation in nine ICE varieties. Although their results indicate that New Englishes (in particular PhilE and HKE) display the strongest preference for the more transparent *of*-genitive (Heller, Szmrecsanyi & Grafmiller 2017: 20, 22), they also showed that all speakers strongly favor

the *s*-genitive with short, animate, given, and thematic possessors (Heller, Szmrecsanyi & Grafmiller 2017: 21). Similar results were obtained by Szmrecsanyi et al. (2016: 123–124), who found that, across all varieties (BrE, CanE, IndE, SgE), possessors realized as proper nouns strongly favor the *s*-genitive, while long possessors are associated with the *of*-genitive. Adopting a production-oriented perspective, Heller, Szmrecsanyi & Grafmiller (2017: 6, 21) related these effects to the principle of “Easy First” (MacDonald 2013: 3), which states that the processing cost during production planning can be reduced by the early use of constituents that can be easily retrieved from memory. Given, animate and highly thematic possessors being easily accessible (Prat-Sala & Branigan 2000; MacDonald 2013: 3), they tend to favor the *s*-genitive, i.e. the variant in which the possessor precedes the possessum.

Similar effects were also found by research on the dative alternation. In constructionist terms, the dative alternation involves the choice between the ditransitive construction (*John gave Mary the book*) and the *to*-caused-motion construction (*John gave the book to Mary*; see e.g. Goldberg 2002: 336), which is also known as the prepositional dative variant. Disregarding a few minor variety-specific effects, such as a stronger preference for the more transparent prepositional dative in IndE (Szmrecsanyi et al. 2016: 130; Röthlisberger, Grafmiller & Szmrecsanyi 2017: 691), or JamE speakers being more likely to use the ditransitive construction with non-pronominal recipients (Röthlisberger, Grafmiller & Szmrecsanyi 2017: 692, 696–698), studies on the dative alternation also suggested that, with respect to processing constraints in L1 and L2 varieties, “there is no [...] fundamental difference between the two types of Englishes” (Bernaisch, Gries & Mukherjee 2014: 27). Studying the dative alternation with the prototypical ditransitive verb *give* in British and South Asian Englishes, Bernaisch, Gries & Mukherjee (2014: 21–22, 26) found that, across all varieties, the ditransitive construction is preferred with pronominal recipients, whereas the prepositional variant is favored with long recipients. Szmrecsanyi et al. (2016: 130), and Röthlisberger, Grafmiller & Szmrecsanyi (2017: 690–691) analyzed a wider range of verbs but also identified the expression type and the length of the theme and the recipient as strongly influencing the constructional choices of speakers of all varieties. According to Röthlisberger, Grafmiller & Szmrecsanyi (2017: 690–691), who analyzed the dative alternation in nine ICE varieties, all speakers tend to use the prepositional dative with long recipients and pronominal themes, while the ditransitive construction is preferred with simple, given, animate and definite recipients. The findings concerning the dative alternation are thus very similar to those obtained for the genitive alternation. The

avoidance of the ditransitive dative with long recipients (i.e. e.g. the use of *John gave the book to his 15-year-old friend who is a great dancer* instead of *John gave his 15-year-old friend who is a great dancer the book*) can again be accounted for by the principle of end-weight or Hawkins' Minimize Domain principle. Ditransitive constructions with long recipient NPs require the processing of additional material before the speaker can parse all constituents of the construction (Hawkins 2004: 26). As a result, this variant is avoided across all varieties. In addition, the association between the ditransitive construction and given, animate and definite recipients indicates that the principle given-before-new (or Easy-First) also plays a major role in the competition between the two dative variants (Röthlisberger, Grafmiller & Szmrecsanyi 2017: 695).

Now, when looking at the overall results of these studies on processing constraints in World Englishes, one central question arises: Why is it that previous research seems to indicate that processing factors have a stronger effect on L2 speakers with respect to preposition placement (Hoffmann 2011), *that*-complementation (Wulff, Lester & Martinez-Garcia 2014; Wulff, Gries & Lester 2018) and, at least to a some extent, particle placement (Gilquin 2015; Paquot, Grafmiller & Szmrecsanyi 2019; Wulff & Gries 2019), whereas no processing-related differences between L1 and L2 speakers have been found with respect to the genitive (Gries & Wulff 2013; Szmrecsanyi et al. 2016; Heller, Szmrecsanyi & Grafmiller 2017; Dubois, Paquot & Szmrecsanyi 2022) and the dative alternation (Bernaisch, Gries & Mukherjee 2014; Szmrecsanyi et al. 2016; Röthlisberger, Grafmiller & Szmrecsanyi 2017)? One reason could be that, compared to preposition placement or *that*-complementation, the dative and the genitive alternation can be said to be less strongly affected by variables that have a direct impact on parsing efficiency (an exception being the effect of constituent length), while e.g. animacy or factors related to information status such as givenness or definiteness play a much more important role (see e.g. Hinrichs & Szmrecsanyi 2007 on the genitive alternation; Bresnan et al. 2007 on the dative alternation). According to Hawkins (2004: 25), “complexity is a function of the number of formal units and conventionally associated properties that need to be processed in domains relevant for their processing”. Consequently, the most efficient construction is “the one that provides the earliest possible access to properties in the ultimate proposition to be communicated” (Hawkins 2004: 25). Although givenness, definiteness and animacy are also related to processing (recall that given, definite and animate constituents can be more easily retrieved from memory), the factors that were found to lead to cross-varietal differences with regard to preposition placement (for more details, see 3.2) and *that*-complementation have a much more direct impact on parsing efficiency

because they increase adjacency and reduce garden paths effects so that fewer elements need to be processed. Stefanowitsch (2003: 439) even argued that the influence of factors related to information status such as givenness “look[s] like a stylistic principle rather than a fundamental cognitive mechanism”. However, for instance, Röthlisberger, Grafmiller & Szmrecsanyi (2017: 695) concluded with respect to the dative alternation “that the prevailing stability in effect direction that we observe across regional varieties of English can be attributed to the principle of easy first”, i.e. to factors such as frequency, givenness and concreteness. Previous research on the role of processing factors in L1 and L2 Englishes thus seems to suggest that processing-related differences between more and less advanced varieties are most likely to emerge with respect to variables that have a direct impact on parsing efficiency. At the same time, factors typically subsumed under the given-before-new principle apparently affect speakers of all varieties in very similar ways.

Another reason that may account for the fact that processing-related differences between varieties have been found in studies on preposition placement, *that*-complementation and particle placement but not with respect to the dative and genitive alternation may be that the former alternation phenomena do not involve major semantic differences, whereas the dative and the genitive constructions are not always semantically interchangeable. While many studies on the dative alternation are based on the assumption that the two dative variants do not differ semantically (see e.g. Larson 1988; Bresnan et al. 2007: 11; Jäschke & Plag 2016: 489), there are also strong arguments against semantic equivalence (see e.g. Oehrle 1976; Gropen et al. 1989: 240–242; Pinker 1989: 97–100; Krifka 2003; for neurolinguistic evidence, see Allen et al. 2012). As pointed out above, Construction Grammar assumes that the two dative variants are separate constructions with different meanings (e.g. Goldberg 1995, 2002; Gries & Stefanowitsch 2004: 102). While the ditransitive construction expresses the meaning ‘X causes Y to receive Z by V-ing’, the *to*-caused-motion construction is associated with the meaning ‘X causes Z to move to Y by V-ing’ (Langacker 1986: 14–16; Goldberg 2002: 332; Hoffmann 2022: 25; for an alternative constructionist view that treats the two constructions as synonymous, see Perek 2015: 154–158). That the two constructions are not completely synonymous is reflected in the fact that certain verbs are strongly associated with certain dative variants. For instance, the ditransitive construction is strongly favored with *give* and other verbs denoting a literal or metaphorical sense of transfer, whereas *bring*, which is fully compatible with the meaning ‘cause something to move somewhere’, is a prototypical verb found in the caused-motion construction (Gries

& Stefanowitsch 2004: 106–107). Moreover, verbs such as *send* or *write* which are semantically compatible with both variants can be used in both constructions, but the resulting constructs are not synonymous (Krifka 2003: 5–6; Gries & Stefanowitsch 2004: 107). For instance, it has been argued that example (2.3a) emphasizes the fact that Joyce ends up being the possessor of the walrus, whereas (2.3b) highlights the path along which the walrus completed its trajectory (examples in (2.3) from Langacker 1986: 14).

- (2.3) (a) *Bill sent Joyce a walrus.*
 (b) *Bill sent a walrus to Joyce.*

Note that I do not mean to suggest that the two dative variants can always be clearly distinguished by a genuine difference in meaning. However, at least in some cases, the choice between the dative constructions may not only be influenced by factors related to parsing, information status or phonology but also by semantic variables since “the specific construal of the events in question will determine the choice of construction” (Gries & Stefanowitsch 2004: 107). Now, of course, one might argue that particularly semantic variables are prone to variation in PCEs (Röthlisberger, Grafmiller & Szmrecsanyi 2017: 698; Szmrecsanyi et al. 2017: 18), and that particularly semantic factors should lead to cross-varietal differences in the choice between competing constructions. However, recall that linguistic innovation in PCEs typically affects the meso-constructural level (cf. 2.2.3.1; also see Bernaisch, Gries & Mukherjee 2014: 26). Cross-varietal differences in constructional choice that can be attributed to semantic factors are therefore unlikely to surface in studies focusing on abstract constructions, such as Szmrecsanyi et al. (2016: 127–133) or Röthlisberger, Grafmiller & Szmrecsanyi (2017). Instead, at a macro-constructural level, the core meanings associated with the two dative variants (i.e. transfer for the ditransitive construction, movement for the caused-motion construction) may lead to relatively similar linguistic choices across all varieties.

Semantic aspects also play a role in the genitive alternation. The *s*-genitive, for example, is generally preferred with prototypical relations between the possessor and the possessum, such as body-part, kinship, ownership or part-whole relations (Rosenbach 2002: 121; Shih et al. 2015: 16). Adopting a constructionist perspective, Stefanowitsch (2003) argued that the two genitive variants constitute two independent semantic-role constructions: “the *s*-genitive assigns the roles POSSESSEE and POSSESSOR to its head and modifier respectively, and the *of*-genitive assigns roles that I have called, for want of a better term, ENTITY and INTRINSIC ENTITY” (Stefanowitsch 2003: 348). Although individual nouns can override the constructions’ semantics, this shows that the meaning

of the two genitive variants is not always identical (Stefanowitsch 2003: 348). However, the investigations of Szmrecsanyi et al. (2016: 122–127) and Heller, Szmrecsanyi & Grafmiller (2017) did not include any variables to account for the effect of the semantic relation encoded by the genitives. The cross-varietal stability these studies found may thus, at least in parts, be due to semantic factors, which may contribute to similar constructional preferences in all varieties.

In contrast, the competition between preposition stranding and pied-piping, between the use and omission of the *that*-complementizer, and between the VPO and VOP constructions does not involve semantic differences. Although certain prepositions are very strongly associated with certain preposition placement variants (Johansson & Geisler 1998: 75; cf. 3.1.2), in contrast to the associations between individual verbs and e.g. the ditransitive dative construction, this effect cannot be attributed to a meaning difference between stranding and pied-piping. *During*, for example, has an almost categorical pied-piping preference, but there is no semantic difference between (2.4a) and (2.4b). (Note, however, that certain stranding and pied-piping constructions can be distinguished by pragmatic differences which will be discussed in 3.1.1.)

- (2.4) (a) *each term **during** **which** negotiations take place* <ICE-CAN:W2D-002#61:1>
 (b) *? each term **which** negotiations take place **during***

With respect to *that*-complementation, speakers tend to omit the complementizer with matrix clause verbs denoting truth claims such as *think* or *believe* (Dor 2005; Kaltenböck 2006: 388). However, this simply means that, depending on the matrix clause verb, *that*-deletion is not always possible. It does not imply that the two variants entail a meaning change, i.e. just like the examples in (2.4), (2.5a) and (2.5b) are semantically interchangeable.

- (2.5) (a) *John believes he will come to the party.*
 (b) *John believes **that** he will come to the party.*

Concerning particle placement, it has been suggested that the VPO construction is preferred with abstract referents of the direct object (2.6a), while the VOP construction is favored with concrete referents (2.7b) (Wulff & Gries 2019: 879). While this seems to be a semantic factor at first sight, it can actually be explained by a processing principle: Since constructions with abstract direct object referents typically involve metaphorical and idiomatic verb-particle combinations (such as *hold up* in (2.6)), these structures favor

the VPO construction (2.6a) because the breaking up of the verb-particle string would lead to an increase in processing cost (Gries 2003: 53–56; Wulff & Gries 2019: 879; examples in (2.6) and (2.7) from Wulff & Gries 2019: 879; own emphasis).

- (2.6) (a) *Charlie **held up** his principles.*
(b) *Charlie **held** his principles **up**.*
- (2.7) (a) *Charlie **held up** the book.*
(b) *Charlie **held** the book **up**.*

In sum, this discussion of previous research on the role of processing constraint in World Englishes therefore arrives at the following conclusion: Processing-related differences between L1 and L2 speakers are most likely to emerge with respect to competing constructions

- (1) whose competition is strongly driven by variables that have an impact on parsing efficiency, and
- (2) which are semantically interchangeable.

Since these two aspects also apply to the competition between preposition stranding and pied-piping (cf. chapter 3), for the present study, it can be expected that L2 speakers should have a stronger preference for constructions that are easy to process than L1 speakers.

2.4.1.2 Psycholinguistic evidence

In addition to the corpus-based studies outlined in the previous section, there is also psycholinguistic research that provides support for the claim that processing constraints have a stronger effect on L2 speakers. Especially important for the present study on filler-gap constructions is the finding that processing-related differences between L1 and L2 speakers are more significant in the processing of long-distance dependencies such as relatives or WH-interrogatives than with regard to local dependencies such as number agreement (Clahsen & Felser 2006a; Dallas & Kaan 2008: 373; Kroll & Dussias 2013). Several reasons have been put forward to account for this effect.

First of all, recall that filler-gap constructions make very high demands on working memory as they involve a lot of simultaneous processing (cf. 1). Now, capacity-based approaches (McDonald 2006; Hopp 2006, 2010) assume that L2 processing is associated with higher cognitive demands than L1 processing due to “less practice in processing the L2 and the effort devoted to the continuous suppression of the L1 in L2

processing“ (Hopp 2009: 466). Particularly under increased processing load, L2 processing has been found to be less efficient and less automatic than L1 processing (McDonald 2006: 397; Hopp 2006: 393, 2010: 926, 2013: 52, 2014: 272). Consequently, it is not surprising that processing-related differences between the two speaker groups are most likely to emerge in complex tasks such as the parsing of long-distance dependencies which require that the filler is held in working memory while other material is being processed (Kaan 2014: 272).

In contrast to capacity-based approaches, which maintain that L2 processing does not differ qualitatively from L1 processing (Hopp 2014: 273), according to Clahsen & Felser’s (2006b: 31–34) “shallow structure hypothesis”, differences in the processing of filler-gap domains (FGDs) between L1 and L2 speakers result from the fact that L2 speakers only rely on lexical information, whereas L1 speakers also make use of syntactic information in language processing. Clahsen & Felser (2006b: 32) thus assume that the syntactic representations that L2 speakers compute are “shallower and less detailed than those of native speakers”. This is particularly relevant in the processing of filler-gap dependencies, in which syntactic information can provide clues for the identification of potential gap sites, thereby facilitating the establishment of a structural relation between the filler and the gap (Marinis et al. 2005: 72; Clahsen & Felser 2006b: 31; Felser & Roberts 2007: 28; Felser et al. 2012: 95).

Other reasons that may explain differences in the processing of filler-gap dependencies between L1 and L2 speakers could be the reduced ability of L2 speakers to use plausibility information (Williams 2006) or to initiate reanalysis after garden path effects (Williams, Möbius & Kim 2001: 532). For the present usage-based study, an interesting approach in this context is Kaan’s (2014: 267–268) view that differences in predictive processing between L1 and L2 speakers can be linked to input. Predictive processing describes “the degree to which information can be pre-activated in anticipation of the input as well as the degree to which prediction errors can be used to adjust future prediction”(Kaan 2014: 261). Now, Kaan (2014: 267) argued that L2 speakers have “weaker associative connections” and “lower-quality lexical representations” than L1 speakers because they are exposed to less input. Due to these inconsistent representations, information retrieval involves more effort for L2 speakers (see Perfetti & Hart 2001), which may explain why these speakers tend to make less predictive use of stored language information than L1 speakers (Kaan 2014: 267–268). Remember that a similar, input-based explanation also forms the basis of the Usage-based Construction Grammar approach to processing outlined in 2.3.1.1, which assumes that L2 processing may

involve higher cognitive cost as the construction networks of L2 speakers are not as deeply entrenched as those of L1 speakers.

In addition, Kaan (2014: 266) suggested that differences in the processing of filler-gap constructions between L1 and L2 speakers may result from the fact that prediction formation is slower in L2 processing since “[b]ilinguals and language learners may activate more information during processing than monolingual native speakers”. Taking a slightly different approach, Cunnings (2017: 662) argued that processing differences between L1 and L2 speakers arise because L2 speakers are more likely to be affected by interference when retrieving information that has been constructed during parsing from memory. As the processing of FGDs requires that the filler is held in memory and activated upon reaching the gap, this aspect may also be particularly relevant in the processing of long-distance dependencies (Cunnings 2017: 661).

Although there is no consensus on the exact ways in which L2 processing differs from L1 processing, research in the field of psycholinguistics indicates that the processing of long-distance dependencies poses a challenge to L2 speakers. Filler-gap constructions being very difficult to process, in the present study on preposition placement, the constructional choices of speakers of all varieties should turn out to be strongly influenced by processing constraints. However, in the light of the psycholinguistic studies outlined in this section, processing factors should have an even stronger impact on L2 speakers of English.

2.4.2 Second-language learning principles

Since PCEs share many characteristics with learner varieties, second-language learning principles constitute another important factor in the emergence of New Englishes. Although the Dynamic Model is not conceptualized as an ENL-ESL classification, according to Brunner and Hoffmann (2020: 3–4),

the model still shows some correlation with it: Stage III is, for all intents and purposes, an assortment of L2 varieties, where usage is marked by effects typical of second-language acquisition. As varieties move into Phase V, in contrast, they reach the level of native languages, where such restrictions play less and less of a role.

Hence, many of the localized linguistic forms that emerge in PCEs can be attributed to second-language learning effects. The term “second-language learning principles” subsumes psycholinguistic principles such as simplification, overgeneralization, regularization or overuse (Bruckmaier 2017: 29) that characterize “the acquisition of any language after the acquisition of the mother tongue” (Ellis 2008: 6). According to Williams (1987: 169, 178), these effects can be summarized by two main principles,

namely “hyperclarity” and “economy of production”, which together reflect the “constant competition between demands for explicitness and demands for economy in language” (Mesthrie & Bhatt 2008: 173). Of particular relevance to the present study are simplification and the “overuse of known forms” (Bruckmaier 2017: 29).

In order to describe the concept of simplification, it is helpful to take a brief look at the notion of complexity. Language complexity generally involves two aspects. On the one hand, there is a quantitative understanding “in structural ‘more is more’ terms” (Steger & Schneider 2012: 156) which simply refers to the amount of language material that is used. From this point of view, simplification describes processes that lead to a reduction in language material, i.e. innovations that result, for example, in shorter overall constructions. One such case of simplification that might be relevant for the present study is the widespread tendency among L2 learners (see e.g. Hulstijn & Marchena 1989; Kao 2001; Liao & Fukuya 2004; Siyanova & Schmitt 2007) to omit “communicatively redundant prepositions” (Gilquin & Granger 2011: 59), which is in accordance with Ellis’ (2006: 19) claim that learners typically fail to use forms that are “redundant in the immediate understanding of an utterance”. On the other hand, complexity also has a qualitative, psycholinguistic dimension (Steger & Schneider 2012: 156). From this point of view, the preference of learners for transparent constructions with one-to-one form-function mappings can also be considered a process of simplification that leads to a reduction in processing effort (Bruckmaier 2017: 30). For instance, Gilquin (2017: 63) argued that the use of *discuss about* instead of *discuss something*, which has been observed in many PCEs (cf. 2.2.3.1), can be attributed to a process of “remapping” (cf. figure 2.4) which aims to “bring[...] syntax and semantics more in line with each other”. According to Gilquin (2017: 61–63), the frequent occurrence of patterns such as *talk/speak/write about* can be assumed to lead to the entrenchment of a [V *about* NP]-construction that is “associated with verbs expressing the conveying of information on a certain subject” (Gilquin 2017: 62). At the same time, speakers can also be expected to store a [V NP]-construction licensing constructs such as *kill somebody* or *break something*, which is associated with the meaning of causation. Now, while speakers of Standard varieties would also use the [V NP]-construction with the verb *discuss* (i.e. *discuss something*), speakers of New Englishes seem to associate *discuss* with the [V *about* NP]-construction, which is more compatible with the prototypical meaning of *discuss*. Similar processes of “remapping” can also account for other constructional

innovations such as *demand for* (by analogy with *ask for*), or *enter into a place* (by analogy with *go/come into*; Gilquin 2017: 63; also see Nesselhauf 2009).²

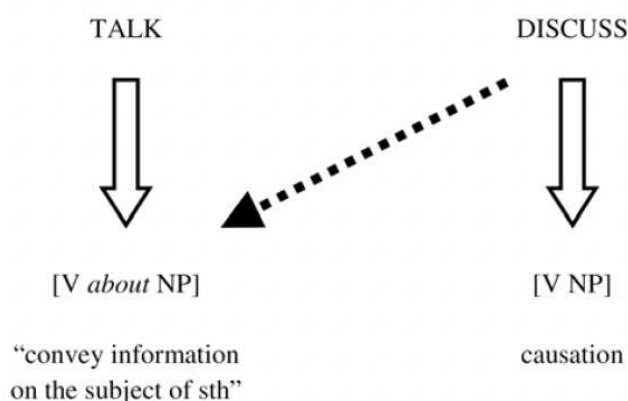


Figure 2.4: The emergence of the [discuss-about]-construction as suggested by Gilquin (2017: 64)

Furthermore, Schneider & Gilquin (2018: 57) observed that learners tend to create more transparent verb/adjective-preposition constructions which include “a semantic, compositional preposition instead of a functional, idiomatic preposition”. In the innovation *indulge into* (instead of *indulge in*), for example, “the preposition iconically reduplicates a directionality instigated by the verb” (Schneider & Gilquin 2018: 67; also see e.g. Gilquin 2015 for the creation of more transparent phrasal verbs). Similarly, Nesselhauf (2009: 20) found that L2 Englishes tend to come up with novel verb-preposition combinations such as *enter into* or *approach to* (instead of *enter/approach*), which she attributed to “a tendency to make the direction expressed in verbs of movement more explicit”. Paradoxically, simplification can thus sometimes even lead to an increase in language material (Mesthrie 2012; Bruckmaier 2017: 30).

A second psycholinguistic principle which causes constructional innovation in L2 Englishes and which might turn out to affect speakers’ constructional choices in this study

² Adopting a different approach, Mukherjee (2007: 175–176; also see Mukherjee & Hoffmann 2006: 166, fn. 16) explained the emergence of *discuss about* based on his concept of “nativized semantico-structural analogy”, i.e. “a process by means of which nonnative speakers of English as a second language introduce new forms and structures into the English language on grounds of semantic and formal templates that already exist in the English language system“. From this point of view, innovations such as *discuss about* or *demand for* are influenced by related nouns that exist in the English language (e.g. *discussion about*, *demand (N) for*; also see Nesselhauf 2009: 19).

is the “systematic overuse [...] of those words that learners can feel safe with”, i.e. of entrenched micro-constructions that speakers feel comfortable using because of their high frequency and their high degree of polysemy (Bruckmaier 2017: 30). Hasselgren (1994: 237) refers to these expressions whose use can be considered a safe strategy as “lexical teddy bears”. For instance, Mukherjee (2009: 131, fn. 9) observed that IndE tends to overuse the high frequency verb *give*, which leads to innovative light-verb constructions such as *give explanation to the problem* (instead of *explain the problem*). An example of overuse involving prepositions comes from Schmied (2004: 934), who found that speakers of East African Englishes choose the high-frequency preposition *in* in cases in which Standard varieties would prefer other, more specific prepositions. This leads to innovative constructs such as *come in the country*, where *in* is used with a directional meaning instead of the more specific *to* (Schmied 2004: 935).

Based on the above, it can be expected that the present corpus study will identify qualitative innovations that can be attributed to second-language learning principles. With respect to the principle of simplification, it can be assumed that speakers do not just omit prepositions, but that PCEs also use innovative preposition placement structures that can be explained by a tendency toward transparency and explicitness. Furthermore, innovations may result from the overuse of prepositions (e.g. *in*) and high-frequency P + WH strings such as *in which* (Jach 2021: 360; cf. 3.1.2).

2.4.3 Cross-linguistic influence

2.4.3.1 Basic terms and assumptions

Another factor that is known to lead to innovations in PCEs is cross-linguistic influence. While a wide range of terms such as “L1 influence”, “interference” or “transfer” are used to describe this phenomenon, in the present study the term “cross-linguistic influence” will be preferred for two reasons: First, compared to the term “L1 influence”, the label “cross-linguistic influence” captures the fact that countries in which PCEs are spoken are typically characterized by multilingual settings in which speakers use a variety of languages on a daily basis (Gut 2011: 104–105). Second, in contrast to the frequently used terms “interference” or “transfer”, the label “cross-linguistic influence” is not as restrictive as it does not only subsume direct transfer of language material but also indirect ways in which languages can influence each other, including e.g. conceptual transfer, avoidance, preference or hypercorrection effects (Gut 2011: 105; Callies 2018: 107; Van Goethem & Hendrikx 2021: 382). In line with Odlin (1989: 27), the present study will thus use the term cross-linguistic influence to describe “the influence resulting from the

similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired”.

While a detailed account of the various ways in which language contact can lead to constructional innovation is beyond the scope of this study (for an overview, see e.g. Gass 1996; Gut 2011), the following examples suffice to show that, from a constructionist view, cross-linguistic influence can manifest itself at all levels of abstraction. An example of micro-constructional innovation resulting from cross-linguistic influence would be the borrowing of specific terms such as the toponymic borrowing that is characteristic of the Foundation phase of the Dynamic Model (see 2.2.3.1). With respect to the meso-constructional level, for instance, Coleman (2016) showed that the emergence of the Dutch ‘time’-*weg* construction (e.g. *We zonnen de middag weg*; Coleman 2016: 100) was influenced by the English ‘time’-*away* construction (e.g. *We sunbathe the afternoon away*). Finally, cross-linguistic influence can also affect macro-constructions, which can, for example, be seen in a study by Schachter (1974), who attributed the avoidance of relative clauses by Japanese and Korean L2 learners of English to the absence of postnominal relative clause constructions in the L1s of these speakers. Due to cross-linguistic influence, English relative clause constructions were not as deeply entrenched in the mental grammars of these learners, which means that they were less likely to be activated.

Since cross-linguistic influence can manifest itself in various ways, it is generally difficult to provide empirical evidence for a direct link between constructional innovation and cross-linguistic influence (Heine & Kuteva 2005: 32; Coleman 2018: 108). Moreover, it can be challenging to disentangle manifestations of cross-linguistic influence from second-language acquisition effects such as simplification or redundancy (for an example of how cross-linguistic influence and L2 acquisition effects interact, see e.g. the use of resumptive pronouns in preposition placement constructions in PCEs; cf. 5.5). Nevertheless, it is generally assumed that the degree of cross-linguistic influence depends on a number of individual, contextual, linguistic and psycholinguistic factors (for an overview, also see e.g. Ellis 2015: 148–149).

Individual factors that have an effect on cross-linguistic influence include, for example, age, personality, motivation or a learner’s aptitude for language learning (Odlin 1989: 129–150; Ellis 2015: 146–147). However, since corpus data is usually not coded for these variables, such subjective aspects cannot be taken into account in the present study (Brunner 2017: 37). A factor that can be controlled for is the proficiency level of speakers, which can be assumed to increase as varieties of English proceed along

Schneider's evolutionary cycle. While it is widely assumed that cross-linguistic influence typically affects the language of low-proficiency learners (Gut 2011: 111), for whom the incorporation of L1 features can serve as a strategy to compensate for their limited L2 resources (Ellis 2008: 343; also see Chan 2004), it can also play a role in the linguistic choices of high-proficiency speakers. For instance, Chan (2004: 66) found that both lower- and upper-intermediate learners show signs of cross-linguistic influence, even though less advanced learners tend to use more L1 features in more complex structures. In addition, Gut (2011: 111) suggested that different proficiency levels are associated with different types of cross-linguistic influence. While direct loans, i.e. the borrowing of micro-constructions, dominate the language of low-proficiency speakers, more advanced learners have been shown to exhibit a greater number of "indirect cross-linguistic influences such as tendencies towards hypercorrection and avoidance" (Gut 2011: 11). From a constructionist view, the finding that cross-linguistic influence also affects the language of advanced learners might be explained by the fact that the construction networks of more proficient speakers are more deeply entrenched than those of less advanced learners (cf. 2.4.1), which might make it easier for more advanced speakers to make cross-linguistic generalizations. With respect to the present study, it can thus be expected that cross-linguistic influence will be a relevant factor in both Phase-III and Phase-IV varieties.

Coming to the contextual factors that influence the degree of cross-linguistic influence, speakers are less likely to be affected by local languages in "focused" settings, in which "learners have a clear idea of what constitutes a language" (Ellis 2015: 141), than in "unfocused" environments, in which speakers do not show much concern for the correct use of linguistic forms. Consequently, cross-linguistic influence should have a greater effect in informal, naturalistic settings than in the context of formal instruction, where speakers tend to focus on the correct production of specific forms (Ellis 2015: 141; also see Odlin 1989). The corpus study in chapter 5 will control for this effect by taking into account the different text types of the ICE corpora.

Furthermore, the degree of cross-linguistic influence also depends on language distance. It is generally assumed that there is a higher probability of cross-linguistic influence if the L1 and the L2 involved are typologically close (Ellis 2015: 135). For instance, Sjöholm (1976) found that the typological closeness between Swedish and Finnish resulted in a learning advantage for Swedish-speaking learners of English, whereas Finnish-speaking learners, whose L1 is not close to English, showed more difficulties in learning. Note, however, that a feature that is similar across languages is

not necessarily easier to learn than a structure for which there is no equivalent in the L1. Instead, the high perceptual saliency of a feature that is very different in the L2 may lead to a “novelty effect” (Gass, Behney & Plonsky 2020: 142; also see Kleinmann 1977), which can make it easier for learners to notice and learn the structure. Hence, it is important to keep in mind that language distance always has to be seen in the context of other factors.

That the relationship between language distance and the degree of cross-linguistic influence is not as simple as frequently assumed was also pointed out by Kellerman (1979, 1983), who introduced the concept of “a learner’s *psychotypology*” (Kellerman 1983: 114; original emphasis). In contrast to actual language distance, psychotypology describes a “learner’s *perception of language distance*” (Kellerman 1983: 114; original emphasis). According to Kellerman (1979, 1983), cross-linguistic influence becomes more likely if a speaker perceives two languages as being closely related, and if the speaker considers a language feature to be “language-neutral” (Kellerman 1979: 40). Language-neutral features are typically frequent and transparent constructions “which a learner believes are common across all languages” (Gass, Behney & Plonsky 2020: 144). In contrast, “language-specific” features (Kellerman 1979: 40) such as the idiomatic construction *kick the bucket* are usually less likely to be affected by cross-linguistic influence as they tend to be perceived as special or marked constructions (Kellerman 1979: 44, 54; Gass, Behney & Plonsky 2020: 147). The relationship of perceived language-specificity and perceived language closeness is also illustrated in figure 2.5. While language-specific constructions are usually only transferred to languages that are perceived as being typologically close, language-neutral features such as the basic declarative clause *The sky is blue* are also transferred to languages that are considered distant (Gass, Behney & Plonsky 2020: 144–147).

	Close						Distant
neutral	X	X	X	X	X	X	X
	X	X	X	X	X	X	
	X	X	X	X	X		
	X	X	X	X			
	X	X	X				
	X	X					
specific	X						

Figure 2.5: Schematized version of Kellerman’s model of language transfer (adapted from Gass, Behney & Plonsky 2020: 146)

Applying Kellerman's (1979, 1983) concept to preposition placement, it can be said that pied-piping is cross-linguistically more common than stranding (Hawkins 1999: 277), which means that stranded constructions should be marked. At the same time, considering that stranded and pied-piped prepositions occur in a wide range of clause types, such as relative and WH-interrogative constructions, which constitute basic filler-gap constructions of languages (cf. 3.1), both preposition stranding and pied-piping can be considered rather language-neutral. Consequently, stranding and pied-piping constructions may even be affected by cross-linguistic influence if speakers do not perceive their local languages and English as being closely related (see Brunner 2017: 41 for a similar line of thinking regarding NP structures).

2.4.3.2 Cognitive approaches to cross-linguistic influence

While the previous section introduced basic factors that must be taken into account in a study that aims to identify effects of cross-linguistic influence, in the following, the concept of cross-linguistic influence will be explained from the perspective of two different cognitive approaches, namely Croft's (2000) theory of language change, and Höder's (2012, 2014, 2018) "Diasystematic Construction Grammar approach". While Croft and Höder conceptualize the mental grammars of multilingual speakers very differently, the most important aspect for the present study is the fact that both approaches assume that cross-linguistic influence starts with the process of "interlingual identification" (Weinreich 1968), i.e. "the mechanism by which linguemes from two different languages in a bilingual individual are identified as the same in some respect" (Croft 2000: 238).

Croft's (2000) theory of language change identifies two main processes of language change, namely "innovation" and "propagation". Innovation refers to the creation of innovative forms through "remappings of the link between form and function" (Croft 2000: 166). In contrast, "propagation" describes the mechanism of social selection that leads to "the establishment of a new convention" (Croft 2000: 95). With respect to cross-linguistic influence, innovation occurs as a result of interlingual identification when speakers establish "a cognitive link between the corresponding linguemes of the two languages with respect to their identity in substance" (Croft 2000: 146). Importantly, this does not mean that corresponding linguemes of different languages share a single representation in the mental grammar of an individual (Croft 2000: 146). Instead, Croft (2000: 147) assumes that each language is associated with a separate lingueme pool. After a speaker has established a cognitive relation between two linguemes from two different

languages, “the counterpart forms in the two languages are rendered variants of a single variable” (Croft 2000: 147). As a consequence, a speaker can then choose between an expression that already exists in the speaker’s L1 and a novel form that belongs to a different lingueme pool but expresses the same function. If this innovative use of a lingueme from a different linguistic code is propagated, this means that “a lingueme [...] has been transferred [...] from the lingueme pool of one language to the lingueme pool of the other language” (Croft 2000: 201).

Interlingual identification also lies at the heart of another cognitive approach to cross-linguistic influence, namely Höder’s (2012, 2014, 2018) “Diasystematic Construction Grammar”, which takes up Weinreich’s (1954) notion of a “diasystem”. Assuming that “there are no a priori language boundaries in the input, in language usage, or in the general cognitive principles that govern the organisation of constructional networks” (Höder 2018: 46), this concept is based on the idea that

[t]he grammatical description of a language system in a multilingual environment – i.e. the socially conventionalised set of all structural elements shared by a specific speaker group as well as cognitively stored and processed by the individual speakers – must include structures of all languages or varieties involved, and the social establishment and individual acquisition of such a system must be inherently multilingual. (Höder 2014: 140)

In contrast to Croft (2000), who maintains that each language is associated with a separate lingueme pool, Höder (2018: 48) thus assumes that speakers do not have separate constructicons for individual languages. Instead, drawing on the constructionist assumption that constructions are associated with specific communicative settings, he claims that speakers have a single multilingual construction network in which language-specificity is only included as part of the pragmatic meaning of constructions (Höder 2018: 44, 48). At the same time, not every construction has to be associated with a specific language, which means that the multilingual construction network contains both language-specific constructions as well as more schematic, language-unspecific “diaconstructions” (Höder 2014: 143).

Although Höder’s diaconstructionist approach offers a useful tool to conceptualize cross-linguistic influence from a constructionist perspective, it is questionable whether speakers’ mental grammars indeed include such maximally abstract, language-unspecific macro-constructions (for studies that do not provide support for a diaconstructionist view, see e.g. Wasserscheidt 2014; Béchet 2020; Koch & Günther 2021). Note that even Höder (2018: 62) himself pointed out that “it is a legitimate question to what extent multilinguals in fact do generalise over the idiosyncrasies in the constructions of their languages”. However, since the present study does not aim to identify language-unspecific preposition placement constructions, for the corpus analysis

in chapter 5, it is only relevant that – despite their different approaches to the mental representations of multilingual speakers – both Croft and Höder emphasize the importance of interlingual identification. As this shows that speakers can be expected to recognize cross-linguistic similarities between linguistic structures, the present investigation will have to take into account potential similarities between preposition placement constructions in English and their counterparts in local languages (cf. section 3.4).

2.5 Summary: The cognitive evolution of PCEs

This chapter presented the evolution of PCEs from a cognitive perspective. After introducing some basic terms (cf. 2.1), it focused on Schneider's (2003, 2007) Dynamic Model (cf. 2.2), which offers a useful theoretical framework for the present study. The Dynamic Model does not just provide a sociolinguistic explanation of the evolution of New Englishes, but, as shown by Hoffmann's (2021) Construction Grammar Approach to the Dynamic Model, it can also be complemented with a cognitive, usage-based perspective that considers input the driving force behind language acquisition. With respect to the construction networks of varieties of English worldwide, Hoffmann's (2021) predictions suggest that varieties at higher stages of evolution should exhibit a higher productivity of constructional slots, whereas less advanced varieties should rely more on prototypical, more substantive preposition stranding and pied-piping constructions. Furthermore, just like the first instances of creative language use in L1 acquisition, qualitative innovations in PCEs should typically affect constructions that contain both substantive and schematic slots.

As pointed out in section 2.3, an input-based explanation can also account for the fact that earlier Dynamic Model phases are frequently characterized by a lack of register distinctions. Since speakers of less advanced varieties primarily encounter formal input, it can be expected that Phase-III varieties rely on preposition placement constructions that are considered formal in Standard varieties even in informal contexts, whereas more advanced varieties should store more pragmatic associations between constructions and informal usage situations.

Furthermore, section 2.4 showed that the corpus study will have to take into account various factors that might contribute to innovations in PCEs. With respect to processing constraints (cf. 2.4.1), two aspects suggest that varieties at lower developmental stages should have a stronger preference for preposition placement

constructions that involve a reduction in processing effort than more advanced varieties: First, previous research indicates that processing-related differences between L1 and L2 speakers are most likely to surface with respect to competing constructions without major semantic differences whose competition is strongly driven by parsing-related variables, i.e. two aspects which, as will be shown in 3.2, also apply to the competition between preposition stranding and pied-piping. Second, stranded and pied-piped prepositions occur in filler-gap constructions, which turned out to be particularly challenging for L2 speakers in psycholinguistic research. Moreover, it can be expected that New Englishes are characterized by constructional innovations that can be attributed to second-language learning principles such as simplification or the overuse of entrenched forms. Finally, differences in the constructional choices between New Englishes and Standard varieties of English may also result from cross-linguistic influence, which can be expected to play a role in both Stage-III and Stage-IV varieties.

While this section provided a cognitive explanation of the evolution of New Englishes, the next chapter will focus on the linguistic phenomenon that will be investigated in the corpus study, i.e. preposition placement in filler-gap constructions.

3. Preposition placement from a usage-based perspective

This chapter aims to point out important factors that affect the competition between preposition stranding and pied-piping. Adopting a usage-based perspective, 3.1 will show which stranded and pied-piped macro- and meso-constructions can be assumed to be entrenched in the BrE construction network. It will be argued that these preposition stranding and pied-piping constructions can be described as a family of preposition placement constructions in which individual constructions are interconnected by both vertical and horizontal constructional relations. Since the present study seeks to identify processing-related differences between varieties, 3.2 will focus on the role of processing factors in the competition between preposition stranding and pied-piping. 3.3 will introduce previous research that pointed out ways in which preposition placement constructions in L2 Englishes may differ from those used in Standard varieties. Finally, 3.4 will provide an overview of the preposition placement constructions that exist in the local languages of the PCEs investigated in the present analysis.

3.1 The family of preposition placement constructions

3.1.1 *The macro-constructional level*

As pointed out in chapter 1, stranded and pied-piped prepositions occur across a wide range of clause types all of which belong to the family of filler-gap constructions. The present study will focus on clause types that allow for both preposition stranding and pied-piping. This includes finite WH-relative (3.1), WH-interrogative (3.2), exclamative (3.3), *it*-cleft (3.4) and topicalization (3.5) constructions (Pullum & Huddleston 2002: 627–628). Although it is widely assumed that free relative clauses (3.6) only license stranding (Pullum & Huddleston 2002: 627–628), Hoffmann (2011: 262) also found instances of free relative constructions with pied-piped prepositions in ICE-GB and ICE-Kenya. The present study will thus also include free relative clause constructions.

- (3.1) (a) *the event **which** he is talking **about***
(b) *the event **about which** he is talking*
- (3.2) (a) ***What** is he talking **about**?*
(b) ***About what** is he talking?*
- (3.3) (a) ***What an event** he is talking **about**!*
(b) ***About what an event** he is talking!*

- (3.4) (a) *It is **the linguistics conference** that he is talking **about**.*
 (b) *It is **about the linguistics conference** that he is talking.*
- (3.5) (a) ***The linguistics conference**, he is talking **about**.*
 (b) ***About the linguistics conference**, he is talking.*
- (3.6) (a) ***What** he is talking **about** is the linguistics conference.*
 (b) *? **About what** he is talking is the linguistics conference.*

While the corpus study will only investigate contexts that license variable preposition placement, a full account of the BrE construction network also requires a look at the construction types with categorical preposition placement preferences. There is a group of clause types that only license preposition stranding, namely *that*-relative (3.7), zero-relative (3.8), comparative (3.9) and so-called “hollow” clause constructions (3.10), i.e. nonfinite constructions (typically *to*-infinitivals) in which a “missing non-subject is [...] recoverable from the matrix clause” (e.g. *The linguistics conference* in (3.10a); Huddleston 2002a: 1175). Prepositional passives such as (3.11) also only allow for stranding (Pullum & Huddleston 2002: 627–628). In contrast, nonfinite WH-relatives (3.12) always require pied-piping (Huddleston, Pullum & Peterson 2002: 1067).

- (3.7) (a) *the event **that** he is talking **about***
 (b) ** the event **about that** he is talking*
- (3.8) (a) ***the event** he is talking **about***
 (b) ** **the event about** he is talking*
- (3.9) (a) ***the same event** as I talked **about***
 (b) ** **about the same event** as I talked*
- (3.10) (a) ***The linguistics conference** is an interesting event to talk **about**.*
 (b) ** **About the linguistics conference** is an interesting event to talk.*
- (3.11) (a) ***The event** has been talked **about** a lot.*
 (b) ** **About the event** has been talked a lot.*
- (3.12) (a) ** This is an interesting event **which** to talk **about***
 (b) *This is an interesting event **about which** to talk.*

Now, of course, all preposition stranding and pied-piping constructions have in common that they involve stranded and pied-piped prepositions. Furthermore, there is no semantic difference between the two variants (e.g. between the (a) and (b) versions in

examples (3.1) - (3.6)). Hence, it could be argued that the phenomenon of preposition stranding and pied-piping can be captured by so-called “allostructions” (Cappelle 2006, 2009), i.e. by “(truth-)semantically equivalent but formally distinct manifestations of a more abstractly represented [preposition placement] construction”. Arguing that it is implausible to assume that speakers do not store “a cognitively real link” (Cappelle 2009: 187) between the VPO construction (*pick up the book*) and the VOP construction (*pick the book up*) in their mental grammars, Cappelle (2006) suggested that the two particle placement constructions are two allostructions of a general verb-particle construction (cf. figure 3.1). Consequently, one might also assume that speakers store a general preposition placement construction that comprises two allostructions, viz. a preposition stranding and a preposition pied-piping construction.

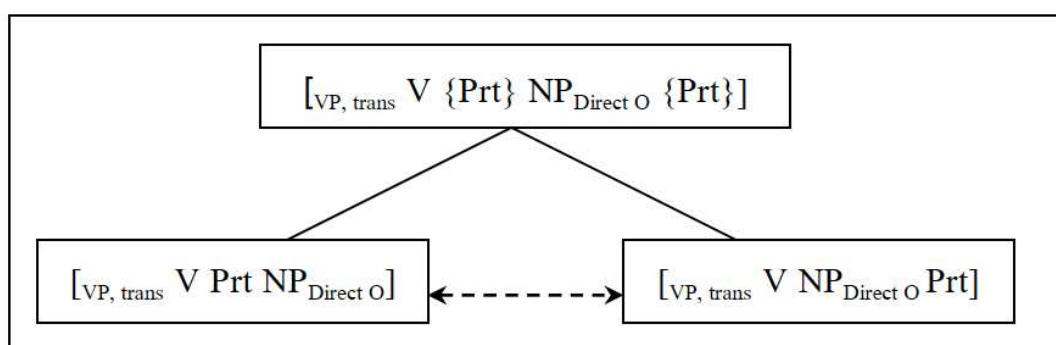


Figure 3.1: The transitive verb-particle construction with its two allostructions (Cappelle 2006: 18)

However, Cappelle’s (2006) concept of allostructions has been criticized for a number of reasons. Hampe (2012: 9) argued that, “from a usage-based perspective, ‘allostructions’ at the most generic level are highly implausible and could [...] be labelled ‘extreme’”. Hoffmann (2020b: 150) pointed out that, due to the fact that an abstract schema such as Cappelle’s general verb-particle construction can never be used to generate concrete utterances such as *pick up the book/pick the book up*, which always rely on more specific constructions (e.g. the VPO or VOP construction), “no data can ever be adduced to falsify [the] existence” of such a superordinate construction. Sommerer & Baumann (2021: 101), who investigated NPN (e.g. *step by step*) constructions in English, are also critical of the value of maximally abstract constructions, arguing that

[i]n a usage-based model which strives for cognitive plausibility, the semi-specific level is an important level one cannot do without. At the same time, the postulation of an abstract overarching schema [...] high up in the network is possible but not really necessary. It is unrealistic that speakers entrench such a general schema in a bottom-up acquisition process [...].

The same kind of criticism thus, of course, also applies to the idea of a general preposition placement construction. Moreover, Hoffmann (2011: 264–275) provided a

detailed, usage-based account of the BrE network of preposition placement constructions, which illustrated that it is impossible to capture all preposition placement phenomena that exist in English by a single abstract construction. As this section will demonstrate, despite superficial similarities, the various constructions in which stranded and pied-piped prepositions occur are so diverse that there is no general preposition placement construction that could unify the constraints of all stranding and pied-piping phenomena. Hence, the term preposition placement “must be considered a theoretical linguistic construct that corresponds to no single construction in a speaker’s mental construction network” (Hoffmann 2011: 264). Even the assumption that speakers store a single abstract preposition stranding and a single abstract preposition pied-piping construction is too general. Instead, the following usage-based account will show that speakers can be said to possess a variety of stranding and pied-piping constructions which can be described as a family of stranding and a family of pied-piping constructions. These two subfamilies form a more general family of interrelated preposition placement constructions.

3.1.1.1 The pied-piping family

Starting with the pied-piping family, a look at the examples in (3.13) - (3.19) (repeated here from (3.1) - (3.6), (3.12)) reveals that, apart from the *it*-cleft in (3.18) and the topicalization in (3.19), all construction types that license pied-piping contain a WH-element.

- (3.13) *the event **about which** he is talking*
- (3.14) ***About what** is he talking?*
- (3.15) ***About what an event** he is talking!*
- (3.16) *? **About what** he is talking is the linguistics conference.*
- (3.17) *This is an interesting event **about which** to talk.*
- (3.18) *It is **about the linguistics conference** that he is talking.*
- (3.19) ***About the linguistics conference**, he is talking.*

Consequently, one might assume that all of the examples in (3.13) - (3.17) can be generated by a single WH-pied-piping construction that can be used to produce pied-piped WH-relatives, WH-interrogatives, exclamatives, free relatives and nonfinite WH-relatives. However, from a Usage-based Construction Grammar perspective, such an assumption is highly implausible.

First of all, Hoffmann's (2011: 156) corpus study showed that in BrE only finite WH-relatives and *it*-cleft constructions are affected by formality, which means that stranding is preferred with informal WH-relatives and *it*-clefts, while pied-piping is favored when these clause types appear in formal situations. The avoidance of stranded prepositions in formal contexts is deeply rooted in English grammar and goes back to 18th-century grammarians, who demanded that the English language take Latin as a model (Aitchison 2001: 9; for more details on this prescriptivist tradition, see Hoffmann 2011: 76–78). Since Latin only permits preposition pied-piping, pied-piped prepositions in English became associated with a “solemn and elevated style”, whereas stranding became restricted to informal contexts such as “common conversation” and “the familiar style of writing” (Lowth 1762; cited in Aitchison 2001: 11). As the following quote from Murray shows, pied-piping was associated with a higher degree of formality not only in WH-relatives but also in WH-interrogatives:

The preposition is often separated from the relative which it governs; as ‘**Whom wilt thou give it to?**’ instead of, ‘**To whom wilt thou give it?**’ [‘]He is an author whom I am much delighted with’; ‘The world is too polite to shock authors with a truth, which generally the booksellers are the first that inform them of.’ This is an idiom to which our language is strongly inclined; it prevails in common conversation, and suits very well with the familiar style in writing: But the placing of the preposition before the relative is more graceful, as well as more perspicuous, and agrees much better with the solemn and elevated style. (Murray 1795: 122; cited in Hoffmann 2011: 76–77; bold emphasis added)

While influential 18th-century grammarians such as Lowth or Murray did not regard preposition stranding as unacceptable per se (Hoffmann 2011: 78), in the nineteenth century, the earlier assumption that stranding should be restricted to informal contexts became known as the “hard-and-fast rule” (Hoffmann 2011: 78) that “[i]t is incorrect to end a sentence with a preposition” (Pullum & Huddleston 2002: 627). As a result, according to Pullum & Huddleston (2002: 627), even today, “older people with traditional educations and outlooks still tend to believe that stranding is always some kind of mistake”. However, present-day English grammars do not reject stranding. Instead, structures with stranded prepositions (3.20b, 3.21b) are associated with informal registers, whereas constructions with pied-piped prepositions (3.20a, 3.21a) are typically preferred in formal contexts (Quirk et al. 1985: 664; Leech 1996: 75; Biber et al. 1999: 107; Trotta 2000: 64; Leech et al. 2009: 233; examples in (3.20) and (3.21) adopted from Quirk et al. 1985: 664; own emphasis).

(3.20) (a) ***At which house** did you leave the car?*

(b) ***Which house** did you leave the car **at**?*

(3.21) (a) *the old house **about which** I was telling you*

- (b) *the old house **which** I was telling you **about***

Importantly, however, the examples in (3.20) and (3.21), taken from the *Comprehensive grammar of the English language*, illustrate that major usage manuals assume that this formality effect applies to both WH-relatives and WH-interrogatives. In contrast, Hoffmann's (2011: 156) corpus study showed that only pied-piped WH-relative and *it*-cleft constructions are associated with a higher degree of formality in BrE, whereas register is not a relevant factor in the competition between stranding and pied-piping in other clause types. Since Construction Grammar assumes that speakers store pragmatic constraints as part of the meaning pole of constructions, the formality effect in WH-relatives and *it*-clefts cannot be accounted for by a general WH-pied-piping construction that also licenses pied-piped WH-interrogatives, exclamatives, free relatives, and nonfinite WH-relatives, which are not affected by formality. In addition to a **general WH-pied-piping construction**, BrE speakers can thus be said to also store a **formal pied-piped finite WH-relative clause construction** and a **formal pied-piped *it*-cleft construction** (as well as an informal stranded finite WH-relative clause construction and an informal stranded cleft construction (cf. 3.1.1.2); see below for reasons why formal pied-piped WH-relatives and *it*-clefts can be considered two separate constructions; Hoffmann 2011: 266).

Furthermore, nonfinite WH-relatives differ from all other clause types due to their obligatory pied-piping requirement (Huddleston, Pullum & Peterson 2002: 1067; Hoffmann 2011: 37), which is illustrated in example (3.22). As speakers should thus not encounter input for stranded nonfinite WH-relatives, from a usage-based perspective, the BrE constructicon should also include a **pied-piped nonfinite WH-relative clause construction** (Sag 1997; Hoffmann 2011: 249).

- (3.22) (a) *if you don't have a base **from which** to choose* <ICE-CAN:S1B-038#47:1:C>
 (b) **if you don't have a base **which** to choose **from***

Finally, there are two clause types that license pied-piping but do not require a WH-element: topicalizations (3.23b), which are also known as "preposed" constructions (Pullum & Huddleston 2002), and *it*-clefts (3.23c), which have already been mentioned above in the context of formality. Both construction types belong to the family of so-called "information structure constructions" as "they deviate in FORM and MEANING from the more frequent and thus in this sense 'more canonical' declarative constructions"

(Hoffmann 2022: 229). For example, compared to the declarative construction in (3.23a), the topicalized construction in (3.23b) and the *it*-cleft construction in (3.23c) emphasize the type of school John went to by presenting this piece of information as new.

- (3.23) (a) *James went to a Spanish school.*
 (b) ***To a Spanish school**, James went.*
 (c) *It is **to a Spanish school** that James went.*

Now, topicalizations are unique in that they license pied-piping but never include a WH-element (Hoffmann 2011: 244). As shown in (3.24a), in pied-piped topicalizations, the filler is always a topicalized PP, whereas the preposed element in the stranded variant is an NP (3.24b). Since pied-piped topicalizations can thus not be generated by any of the constructions postulated so far, speakers can also be assumed to have a **pied-piped topicalization construction**.

- (3.24) (a) [***To Ferndale***]_{i-PP-top}, *I wrote*_i.
 (b) [*Ferndale*]_{i-NP-top}, *I wrote* **to**_i <ICE-GB:S1B-064 #98:1:B>

Just like in topicalizations, the filler in *it*-clefts is also a topicalized NP (in the case of stranding; (3.25a, 3.26a)) or PP (in the case of pied-piping; (3.25b, 3.26b)). This topicalized element is preceded by *it* and a form of *be* (such as *it is/it was*), and followed by “a relative-like dependent clause” (Biber et al. 1999: 959). Hence, *it*-clefts do not only share properties with topicalizations, but, due to the presence of a WH- or *that*-clause, they are also superficially similar to relative clause constructions. However, while *that*-relatives always require preposition stranding, *it*-clefts with *that* also license pied-piping (3.25b). This can be explained by the fact that, in contrast to relatives, the WH-/*that*-element in *it*-clefts does not function as a relativizer. In (3.25) and (3.26), for example, the WH-/*that*-clauses that make up the second part of the cleft constructions do not postmodify the NP *those patients* (Ward, Birner & Huddleston 2002: 1416; Hoffmann 2011: 255–256).

- (3.25) (a) *it's* [***those patients***]_{i-NP-top} [***that I'll be talking about*** _i]. <ICE-GB:S2A-029 #14:1:A>
 (b) *it's* [***about those patients***]_{i-PP-top} [***that I'll be talking*** _i]
 (3.26) (a) *it's* [***those patients***]_{i-NP-top} [***who I'll be talking about*** _i].
 (b) *it's* [***about those patients***]_{i-PP-top} [***who I'll be talking*** _i].

The fact that *it*-clefts exhibit characteristics of both relatives and topicalizations suggests that speakers have an independent pied-piped *it*-cleft construction. Taking into account the formality effect described above, speakers should possess a formal pied-piped *it*-cleft construction as well as a formal pied-piped WH-relative clause construction.

Note, however, that, in addition to the (a) and (b) variants in examples (3.25) and (3.26), there is a third construction type whose structure is more comparable to that of traditional relative clauses. In example (3.27a), in which the topicalized element is an NP and the gap is a PP, the relative clause can indeed be considered to postmodify the NP *those patients* (Hoffmann 2011: 256). Furthermore, just like a *that*-relative construction, the *that*-example in (3.27b) does not license pied-piping. Following Hoffmann (2011: 114), in the present study, such cases will thus not be considered *it*-clefts but pied-piped WH-relatives.

- (3.27) (a) *it's* [*those patients*]_{NP-top} [[*about whom*]_{PP-i} *I'll be talking*]_i]_{RC}.
 (b) * *it's* [*those patients*]_{NP-top} [[*about that*]_{PP-i} *I'll be talking*]_i]_{RC}.

This usage-based account of preposition pied-piping has shown that speakers can be said to store at least five abstract pied-piping constructions:

- a general WH-pied-piping construction,
- a pied-piped nonfinite WH-relative clause construction,
- a formal pied-piped finite WH-relative clause construction
- a formal pied-piped *it*-cleft construction, and
- a pied-piped topicalization construction.

Crucially, though, the fact that there is not a single superordinate pied-piping construction does not imply that speakers do not store associations between similar constructions in their mental grammars. However, as Hoffmann (2020) pointed out, instead of postulating maximally abstract constructions (such as the allostructions suggested by Cappelle 2006, 2009; cf. 3.1.1), “in many cases [knowledge of language] might be better analyzed as **a link in the network**” (Hoffmann 2020b: 151; original emphasis). While the constructionist literature has so far mostly focused on vertical relations between constructions, it can be assumed that “the grammar network [also] involves horizontal, or lateral, relations between semantically or formally similar constructions at the same level of abstraction” (Diessel 2019: 199; for more on horizontal constructional relations, also see Audring 2019; Lorenz 2020; Smirnova & Sommerer 2020; Zehentner & Traugott 2020; Sommerer & Baumann 2021). Drawing on evidence

from research on L1 acquisition, language change and structural priming, Diessel (2019: 222) argued that “semantically and/or formally similar constructions [...] are stored and processed in an activation network similar to lexemes in the mental lexicon”. For instance, with respect to structural priming, Diessel (2019: 204) pointed out that

[t]here are conspicuous parallels between lexical and structural priming, supporting the hypothesis that constructions are stored and processed in an association network. Like lexical access, access to constructions involves competition between similar constructions that is influenced by the activation level of the competing constructions at a particular moment in time. The activation level is contingent on the degree of entrenchment and priming. Like lexemes, constructions are easily activated when they are deeply entrenched in memory or primed by the same or a similar element in the previous discourse, suggesting that speakers’ grammatical knowledge involves associative connections between structurally and/or semantically similar grammatical patterns that influence constructional access through activation spreading.

In a similar vein, Van de Velde (2014: 147) argued that, in addition to the vertical dimension of constructional networks,

[t]here is [...] another respect in which constructions form a network: rather than just forming a hierarchical structure, constructions can also be related to each other on what could be called the horizontal axis. What I have in mind here is a network where the form-function relation of a particular construction may be partly motivated in relation to its neighbours.

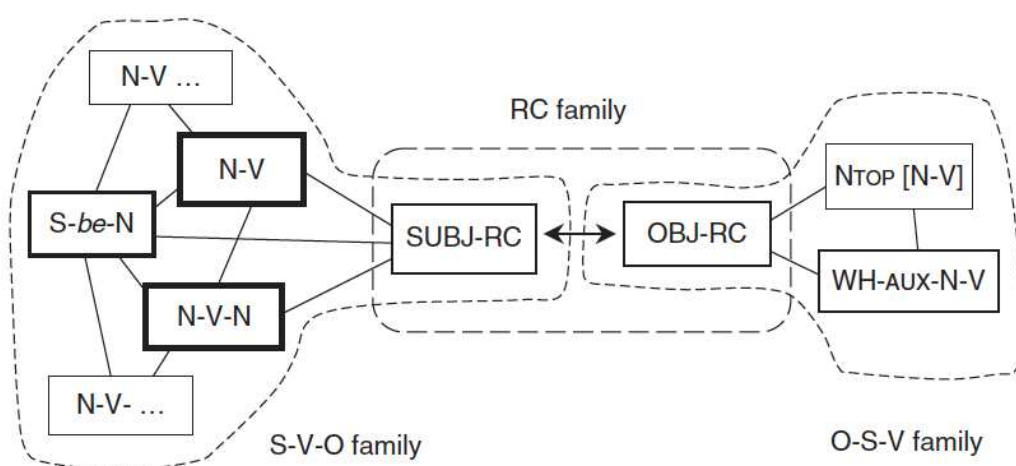


Figure 3.2: Relationship between subject and object relative clauses and their construction neighbors (Diessel 2019: 208)

Van de Velde (2014: 149–150) applied the concept of horizontal constructional relations to account for the fact that Dutch (like German) has three options of placing finite verbs: the clause-initial (“V₁”), the second (“V₂”), and the clause-final position (“V_n”). According to Van de Velde (2014: 150), the V₁-, V₂-, and V_n-constructions are not stored as isolated items, but they are linked by horizontal relations. Adopting a similar approach, Diessel (2019: 207–208) treated the different types of relative clauses in English as a family of interconnected relative clause constructions in which e.g. subject and object relative constructions are linked by horizontal relations. In addition, individual

relative clause constructions are also connected to “sentence neighbors” (Wells et al. 2009; cited in Diessel 2019: 207). For example, the link between subject relatives and the S-V-O family in figure 3.2 indicates that subject relative constructions (e.g. *the boy who won the competition*) share many characteristics with declarative main clauses. Object relative clauses (e.g. *the competition which the boy won*) are also linked to neighbors, namely topicalization (e.g. *The competition, the boy won*) and WH-interrogative constructions (e.g. *Which competition did the boy win?*), but they generally have fewer sentence neighbors than subject relatives as the O-S-V family is smaller than the S-V-O family (Diessel 2019: 207–208).

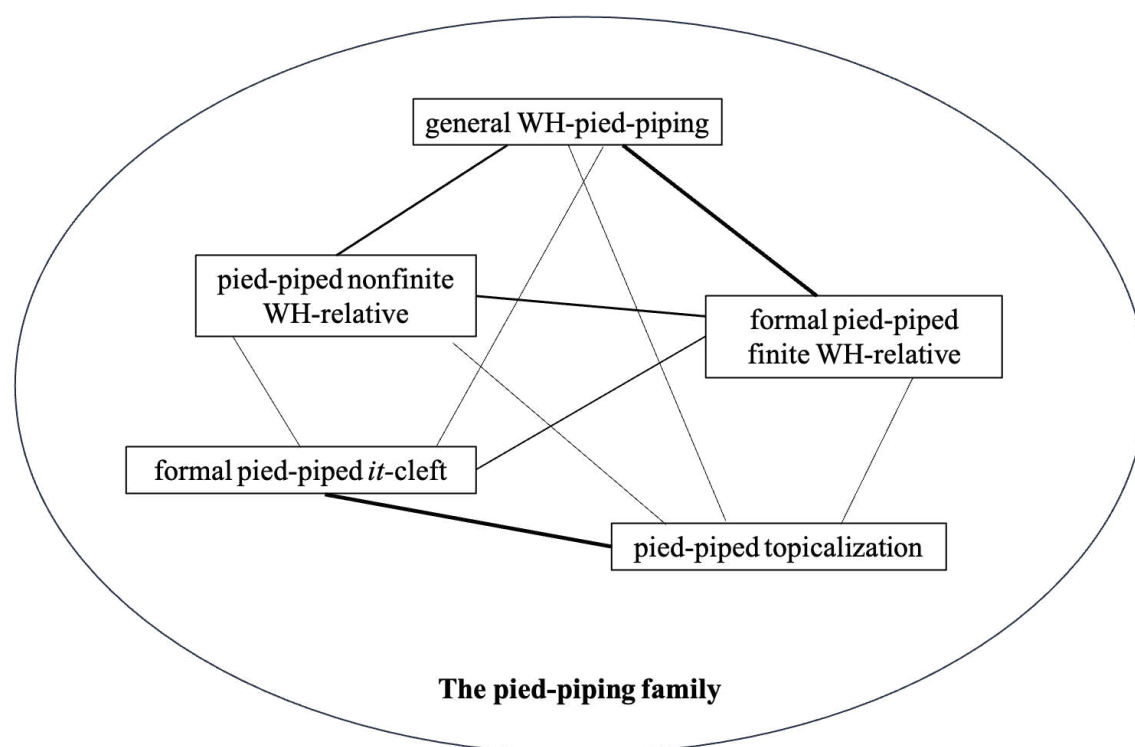


Figure 3.3: The family of pied-piping constructions in BrE

Following this concept of horizontal constructional relations, I suggest that the various pied-piping constructions that can be assumed to be entrenched in the BrE construction network can be described as a family of interconnected pied-piping constructions (cf. figure 3.3). All pied-piping constructions that form part of the circle in figure 3.3 can be said to be horizontally connected at the same level of abstraction. The thickness of the linking lines indicates the degree of association between the individual constructions. For instance, the three constructions that include a WH-element should be more strongly interconnected than e.g. the general WH-pied-piping and the pied-piped topicalization construction since the latter does not include a WH-element. At the same

time, there is also a strong link between the formal pied-piped *it*-cleft construction and the pied-piped topicalization construction, which both belong to the family of information structure constructions. The relationship between the pied-piping constructions and other sentence neighbors will be explained in 3.1.1.3.

3.1.1.2 The stranding family

The various stranding phenomena that exist in English cannot be unified by a single abstract construction either. That speakers can be said to store informal counterparts to the formal pied-piped finite WH-relative clause construction and the formal pied-piped *it*-cleft construction – i.e. an **informal stranded finite WH-relative clause construction** and an **informal stranded *it*-cleft construction** – has already been pointed out in 3.1.1.1. If speakers had only one other general stranding construction, this construction would have to license all of the following constructs (examples (3.28) – (3.36) repeated here from (3.2) – (3.3) and (3.5) - (3.11)):

- (3.28) *What is he talking about?*
- (3.29) *What an event he is talking about!*
- (3.30) *The linguistics conference, he is talking about.*
- (3.31) *What he is talking about is the linguistics conference.*
- (3.32) *the event that he is talking about*
- (3.33) *the event he is talking about*
- (3.34) *the same event as I talked about*
- (3.35) *The linguistics conference is an interesting event to talk about.*
- (3.36) *The event has been talked about a lot.*

Now, in theory, a general stranding construction could indeed be used to produce stranded WH-interrogatives (3.28), exclamatives (3.29), topicalizations (3.30), free relatives (3.31), *that*-relatives (3.32), zero-relatives (3.33), comparatives (3.34) and hollow clauses (3.35) (Hoffmann 2011: 263). For example, the construct in (3.32) could be generated by combining a general stranding construction with a *that*-relative clause construction. However, from a usage-based perspective, it seems more plausible that speakers store additional lower-level stranding constructions.

First of all, as shown in (3.37) and (3.38), both *that*- and zero-relatives only license preposition stranding (Quirk et al. 1985: 1254; Hoffmann 2011: 37).

- (3.37) (a) *the world that I was working in* <ICE-GB:S1A-001 #35:1:B>

- (b) * *the world **in that** I was working*
- (3.38) (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*

Interestingly, this constraint also applies to nonfinite zero-relatives (3.39) even though nonfinite WH-relatives only allow for pied-piping (cf. example (3.22); Hoffmann 2011: 37). Note that nonfinite relatives do not permit the use of a *that*-relativizer (Huddleston, Pullum & Peterson 2002: 1057).

- (3.39) (a) *it would be a great thing **Ø** to put his name **to*** <ICE-GB:S1A-003
#55:1:B>
- (b) * *it would be a great thing **to Ø** to put his name*

With respect to the present discussion of the BrE construction network, this suggests that speakers should receive a lot of input for zero- and *that*-relative clause constructions with stranded prepositions, whereas they should not be exposed to pied-piped non-WH-relatives. Since input drives entrenchment (cf. 1.2.1), the strong association between *that*-/zero-relatives and preposition stranding can be expected to lead to the entrenchment of a **stranded *that*-relative clause construction** and a **stranded zero-relative clause construction**.

While *that*- and zero-relatives are categorical stranding environments, WH-interrogatives, exclamatives, and free relatives allow for both stranding and pied-piping. Nevertheless, stranding can be considered the prototypical choice in all three clause types. In WH-interrogatives, the stranding preference can be explained by the discourse function of the WH-word. According to Trotta (2000: 55),

[s]emantically and functionally, the interrogative wh-word represents unspecified information which characteristically has not previously been introduced into the discourse. In the typical communicative function of interrogatives as questions, it is the wh-word which signals interrogation and should logically come early to successfully fulfill that purpose.

In interrogatives, stranding is thus strongly favored over pied-piping as the former variant ensures the clause-initial placement of the WH-element (Trotta 2000: 55, 107; Hoffmann 2011: 155, 268). The usage-based construction network of BrE should therefore include a **stranded WH-interrogative construction** (Hoffmann 2011: 266).

Likewise, Trotta (2000: 107) claims that the WH-word assumes an important functional role in exclamatives:

[T]he form of this clause type is crucially dependent on placing the element that shows surprise, dismay, excitement, etc, in the thematic ‘slot’ of the sentence, which is clause initial in English. The chunk of that fronted element whose specific function it is to signal exclamation is the wh-

word, and, as in interrogatives, it must come as early as possible within that element in order to successfully realize the intended illocutionary force of the exclamative.

Consequently, stranding should also be preferred in exclamatives (Quirk et al. 1985: 834). However, exclamatives being a rather infrequent construction type (Trotta 2000: 101), it is doubtful whether the limited amount of input speakers are exposed to indeed leads to the entrenchment of an abstract stranded exclamative construction. Trotta (2000: 101) observed that exclamatives make up less than 1% of all WH-tokens in the Brown corpus, while Hoffmann (2011: 136) found only one stranded exclamative token in the ICE-GB. In order to shed more light on speakers' cognitive representations of stranded exclamatives, this phenomenon needs to be analyzed with larger sample sizes.

Stranding is also strongly favored in free relatives. As pointed out in 3.1.1, free relatives are frequently treated as categorical stranding environments (Pullum & Huddleston 2002: 627–628), which has to do with the fact that the WH-element in free relatives is a combination of the antecedent and the relativizer. In (3.40), for example, *who* is equivalent to the ANTECEDENT + WH string *the person who* (*I played with*). Consequently, a pied-piped preposition would precede both the antecedent and the relativizer (* *with the person who I played*), which is considered ungrammatical in English (Pullum & Huddleston 2002: 628–629; Hoffmann 2011: 38).

(3.40) (a) *Well that's **who** I played **with** over Christmas in the Maltings in Aldeborough* <ICE-GB:S1A-058 #148:2:A>

(b) *? That's **with whom** I played over Christmas.*

Hoffmann (2011: 262) speculated that pied-piping in free relatives is only possible when the form of the filler of the free relative clause is compatible with the requirements of the verb of the matrix clause. From this point of view, pied-piping in (3.40b) is ungrammatical because the filler of the free relative clause is a PP (*with whom*), while the main verb (*be*) of the matrix clause requires an NP. At the same time, such an explanation would account for the fact that pied-piping seems to be possible in the following example (3.41), which surfaced in Hoffmann's (2011: 120) ICE-GB data:

(3.41) *the Dutch have been exporting Edam cheese in large quantities to Germany but via such exotic routes as Andorra in the Pyrenees and Tanzania **in whichever country that lies*** <ICE-GB:S1A-061 #325:1:B>

According to Hoffmann (2011: 262), in (3.41), “the free relative clause functions as an optional adverbial. Consequently, the main verb does not impose any requirements on the form of the free relativizer, which results in pied-piping [...] being grammatical”.

However, even though this suggests that – contrary to the assumptions of major English grammars such as the *Cambridge grammar of the English language* – pied-piping is not totally impossible in free relatives, the stranded variant can be expected to surface much more frequently in speakers’ input (for the distribution of stranded and pied-piped free relative tokens in Hoffmann’s (2011) ICE-GB data, see table 3.1). As a consequence, BrE speakers should have an entrenched **stranded free relative clause construction** (Hoffmann 2011: 266).

The remaining clause types that still need to be accounted for are stranded topicalization, comparative and hollow clauses as well as prepositional passives. While topicalizations allow for both stranding and pied-piping, compared to e.g. WH-interrogatives or free relatives, which strongly favor stranding, topicalizations are not known to have a preference for either preposition placement variant (Hoffmann 2011: 55). Instead, the choice between stranded (3.42b) and pied-piped (3.42a) topicalizations seems to be largely determined by context since topicalized elements “must be discourse-old, [i.e.] acting as a link to other entities evoked in the prior discourse” (Ward, Birner & Huddleston 2002: 1373). Moreover, according to Hoffmann (2011: 55), the stranded and the pied-piped variant emphasize different kinds of information. For example, (3.42a) highlights the fact that the person wrote to Ferndale (i.e. e.g. (s)he did not talk or go to Ferndale), while (3.42b) focuses on the entity to whom the writing was addressed (i.e. to Ferndale but no to another company).

- (3.42) (a) [*To Ferndale*]_{i-PP-top}, *I wrote*_i.
 (b) [*Ferndale*]_{i-NP-top}, *I wrote* **to**_i <ICE-GB:S1B-064 #98:1:B>

Furthermore, it is important to note that topicalizations are a rather infrequent construction type as all information structure constructions “have a syntactically more basic counterpart differing [...] in the way the informational content is presented” (Ward, Birner & Huddleston 2002: 1365). Since these “more basic” constructions (e.g. the declarative clause *I wrote to Ferndale*) can be considered the more prototypical variants, these canonical structures should be used much more frequently than topicalizations such as (3.42a) or (3.42b). As speakers can thus be expected to encounter stranded topicalizations with a low type frequency, this structure is most likely not represented as an independent construction in the BrE construction network.³

³ Note that this assumption is compatible with the claim made in 3.1.1.1 that speakers possess a pied-piped topicalization construction. Although pied-piped topicalizations can also be considered an infrequent

Table 3.1: ICE-GB results for categorical clause contexts (Hoffmann 2011: 120)

Clause type	Stranded	Pied-piped
Finite non-WH-relative clauses	348	0
Free relative clauses	157	4
Passive clauses	97	0
Hollow clauses	14	0
Comparison	3	0
Total	619	4

Comparatives and hollow clauses are also low-frequency phenomena. As shown in table 3.1, Hoffmann (2011: 120) found only 14 stranded hollow and three stranded comparison tokens in the ICE-GB. This indicates that stranded hollow clauses and stranded comparatives surface with considerably lower frequencies in speakers' input than e.g. stranded non-WH-relatives or stranded free relatives. Just like stranded topicalizations, stranded comparatives and stranded hollow clauses are thus unlikely to be represented as independent constructions in the network of preposition placement constructions. Instead, speakers can construct these structures by combining other constructions. For instance, stranded topicalizations can be generated out of the general stranding construction and a general topicalization construction in which an NP is topicalized (see e.g. Hoffmann 2022: 228). However, as this involves more processing cost than relying on entrenched constructions, stranded topicalizations, comparatives and hollow clauses can be expected to occur less frequently in actual language use than e.g. stranded WH-interrogatives or free relatives, for which speakers have entrenched constructions (Hoffmann 2011: 266-267).

Finally, the prepositional passive differs from all other stranding contexts as it is not associated with a specific clause type. Instead, constructions in the passive generally only license stranding (Hoffmann 2011: 257). Hence, for instance, the finite WH-relative clause construction in (3.43), in which the subject also functions as the prepositional complement, only allows for stranding:

- (3.43) (a) *The trade war is an artificial barrier **which** must be dealt **with***
 <ICE-CAN:W2E-008#18:1>
- (b) * *The trade war is an artificial barrier **with which** must be dealt*

construction type, pied-piped topicalizations are the only structures that license pied-piping but never contain a WH-element. Hence, speakers should have a separate construction to account for this phenomenon.

Since prepositional passives are a categorical stranding environment and thus not relevant for the present analysis, a detailed discussion of this phenomenon is beyond the scope of this study (for an overview, see e.g. Quirk et al. 1985: 1155; Ward, Birner & Huddleston 2002: 1433–1434; Brinton & Traugott 2005: 128; Schwarz 2019). Nevertheless, the idiosyncratic constraints associated with stranded passives must be accounted for by a **prepositional passive construction**. The meaning pole of this constructional template specifies that the prepositional passive construction involves a patient that is affected by the forces of an agent (Bolinger 1977: 67; Hoffmann 2022: 205–206). On the formal level, the NP appearing in the subject slot also acts as the prepositional complement of the stranded preposition. Moreover, prepositional passives differ from other categorical clause contexts such as hollow or comparative clauses as they license stranding only with two types of prepositions, viz. (the following list is taken from Hoffmann 2011: 52; also see Ward, Birner & Huddleston 2002: 1433):

- (1) prepositions which are lexically specified by verb
- (2) locative prepositions which are not specified by a verb

For instance, in (3.44), preposition stranding is possible since *approve* must be followed by *of*. The idiom *lose sight of* in (3.45) is also stored as a lexicalized item. In contrast, the locative preposition *in* in (3.46) is not specified by *slept* (examples taken from Ward, Birner & Huddleston 2002: 1433).

(3.44) *The plan was **approved of** by my mother.*

(3.45) *The main goal seems to have been **lost sight of**.*

(3.46) *This bed has been **slept in**.*

As these constraints only apply to stranding with prepositional passives but none of the other construction types discussed above, speakers must also store a prepositional passive construction.

In sum, this usage-based discussion of preposition stranding has shown that speakers of BrE can be said to possess several clause-type specific stranding constructions as well as a slightly more abstract general stranding construction which can be used to generate infrequent types such as stranded comparative or hollow clauses. Just like the pied-piping constructions (cf. 3.1.1.1), all these stranding constructions can be treated as an interconnected family of stranding constructions, which is visualized in figure 3.4.

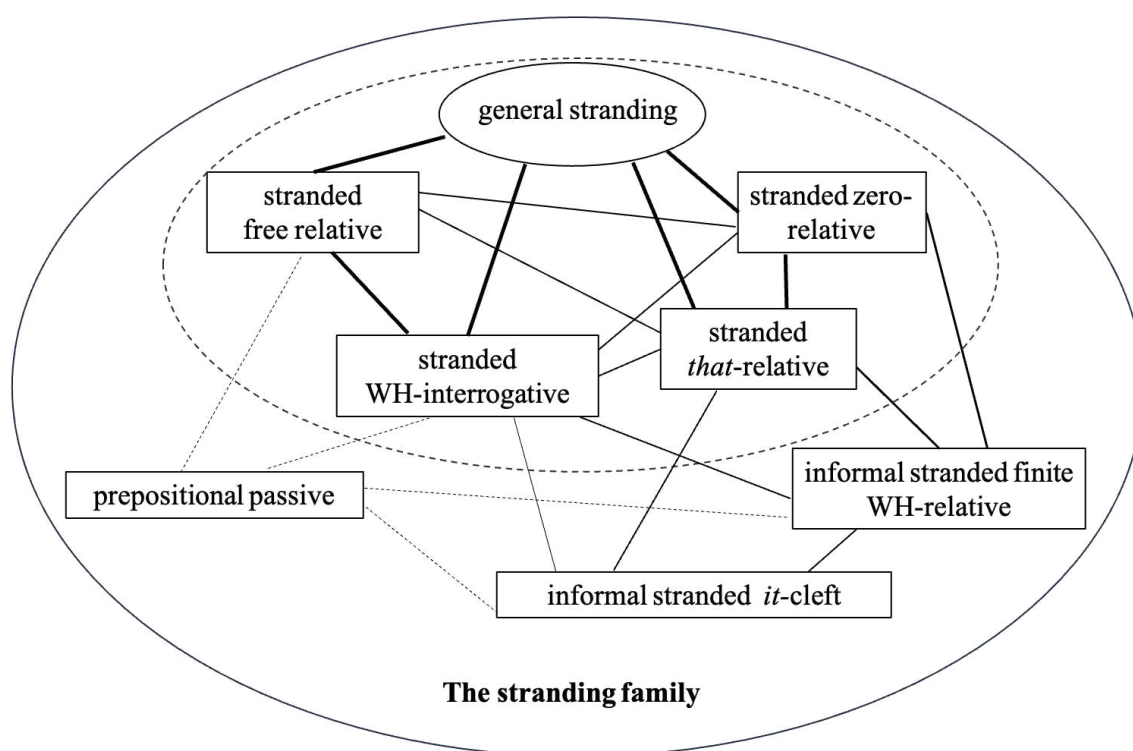


Figure 3.4: The family of stranding constructions in BrE

Recall that the general stranding construction could, in theory, also license stranded WH-interrogative, free relative, *that*-relative and zero-relative constructions. However, as outlined above, it is cognitively more plausible that speakers store additional lower-level constructions. The circle marking the general stranding construction in figure 3.4 thus indicates that the general stranding construction can be considered slightly more abstract than the other constructions. All other constructions are linked to each other by horizontal relations at the same level of abstraction. The thickness of the lines again reflects the strength of the constructional relations. For instance, the stranded zero-relative and the stranded *that*-relative construction can be said to be very strongly connected since they differ only with respect to the type of the relativizer. Likewise, there is a strong link between the stranded free relative construction and the stranded WH-interrogative construction. Not only do both types contain a WH-element but, as shown in (3.47), free relatives and embedded interrogatives can also be ambiguous. Although there is a set of formal and semantic criteria that can be applied to distinguish the two construction types (for an overview, see Quirk et al. 1985: 1050–1053, 1056–1059; Huddleston, Pullum & Peterson 2002: 1070–1073), the only difference between the constructs in (3.47a) and (3.47b), for example, is that the embedded interrogative (3.47b) “leave[s] a gap of unknown information” (Quirk et al. 1985: 1051; i.e. it contains the question *What did she write?*), whereas in the free relative example (3.47a) the unknown

element can be identified through the antecedent contained in the filler (i.e. in (3.47a) *what* can be paraphrased by *the thing which*; Huddleston, Pullum & Peterson 2002: 1070–1071). At the same time, (3.47c) is ambiguous as the subject can be interpreted as an embedded interrogative or as a free relative. With an embedded interrogative interpretation, the meaning of (3.47c) is ‘it is unknown what kind of thing she wrote’. However, if *what she wrote* is taken as a free relative, the overall interpretation of the construction is ‘the thing that she wrote is completely unclear because of the bad writing style’ (examples in (3.47) adapted from Huddleston, Pullum & Peterson 2002: 1070).

- (3.47) (a) *I really liked **what she wrote**.*
 (b) *I can't help wondering **what she wrote**.*
 (c) ***What she wrote** is completely unclear.*

As this shows that free relatives and embedded interrogatives can frequently only be distinguished by subtle semantic differences, the two constructions should be strongly interconnected in speakers' mental grammars. Furthermore, the thin linking line between the prepositional passive and the other constructions indicates that, due to the idiosyncratic constraints of stranded passives, the prepositional passive construction is only very loosely connected to the other stranding constructions. Finally, note that, for reasons of space, figure 3.4 does not include all possible horizontal links between constructions.

3.1.1.3 The preposition placement family

Based on the assumption that speakers store links between similar stranding and between similar pied-piping constructions, the previous sections introduced a family of stranding (cf. 3.1.1.1) and a family of pied-piping constructions (cf. 3.1.1.2). However, a cognitively plausible account of the construction network additionally needs to find a way to account for the fact that speakers can also be expected to establish relations between stranding and pied-piping constructions. For instance, speakers should not just recognize similarities between e.g. (3.50a) and (3.51a), or between (3.50b) and (3.51b), but between all the examples in (3.48) - (3.51):

- (3.48) *I think this is not going to be done in the short time **that you're talking about** <ICE-GB:S1A-024 #78:1:A>*
 (3.49) *And **what I'm obviously talking about** here is the classic phenomena of bureaucracy <ICE-GB:S2A-049 #64:1:A>*

- (3.50) (a) *the Chancellor still has to leave room for the kind of income tax cuts **which the Prime Minister keeps talking about*** <ICE-GB:S1B-052 #20:1:D>
- (b) *The speaker values the person **to whom he is talking*** <ICE-IRE:W1A-012\$A>
- (3.51) (a) *now **what's he talking about*** <ICE-GB:S2A-050 #59:1:A>
- (b) *you probably know **to whom you're talking*** <ICE-GB:S2A-028 #109:2:A>

Support for the claim that all these constructions are somehow related in speakers' mental grammars comes from diachronic research. Historically, the stranded WH-relative construction and the stranded WH-interrogative construction emerged as a result of analogical extension of the stranded *that*-relative construction and, possibly, also of stranded free relative constructions.

In Old English (OE), WH-interrogatives (3.52) and WH-relatives (3.53) only licensed pied-piping (Allen 1980: 270, 284; Bergh & Seppänen 2000: 295). Alternatively, OE relative constructions could also be constructed with the indeclinable relativizer *ðe* (3.54), which corresponds to *that* in present-day English. In contrast to WH-relatives, these *ðe*-relative constructions only permitted preposition stranding (Allen 1980: 267).

- (3.52) *Ic nat ful geare **ymb hwæt** ðu giet tweost.*

‘I do not know full well about what you still doubt.’

(Boeth.V.3 p.12.26; cited in Allen 1980: 285; own emphasis)

- (3.53) *Gehyr ðu arfæsta God mine stefne, **mid ðære** ice arm to ðe cleopie.*

‘Hear thou, merciful God, my voice, with which I, poor one, cry to thee.’

(Blickling p.89.13; cited in Allen 1980: 270; own emphasis)

- (3.54) *. . . on ðære ylcan byrig **ðe** crist **on** ðrowode*

‘in the same city that Christ suffered in’

(ALC.S.XXVII.9; cited in Allen 1980: 268; own emphasis)

However, in the 13th century, stranded prepositions also started to appear in WH-relatives (3.55), and later also in WH-interrogatives (3.56). As a result, in Late Middle English, WH-relative and WH-interrogative constructions were no longer associated with an obligatory pied-piping requirement (Bergh & Seppänen 2000: 297–299; Fischer et al. 2000: 84–85).

(3.55) *That Fortune wolde that he moste twynne/ Out of that place, **which** that I was **inne***

‘That Fortune would have it that he must depart out that place which I was in’

(Chaucer, *The Squire’s Tale*, c. 1395; cited in Bergh & Seppänen 2000: 298; original emphasis)

(3.56) *nuste nan kempe, **whæm** he sculde slæn **on***

‘No soldier knew whom he should strike at’

(*Brut*(Clg) 13718-19; cited in Fischer 1992: 390; own emphasis)

Two aspects are likely to have contributed to this extension of stranding to WH-interrogatives and WH-relatives. First, speakers might have started to use stranded prepositions in *which*-relatives due to the fact that both *that* and *which* are indeclinable and thus easily interchangeable (Fischer 1992: 391; also see Hoffmann 2011: 63). Second, according to Fischer (1992: 297–300, 391), in ME, new WH-relatives emerged out of OE free relative constructions, which licensed preposition stranding at least in some cases. As embedded interrogatives constructions such as (3.56) share certain properties with free relative constructions (cf. 3.1.1.2), it is not surprising that stranding became an option in WH-interrogatives as well (Fischer 1992: 391). The analogical extension of preposition stranding to WH-interrogatives and WH-relatives in ME thus provides support for the claim that speakers recognize similarities and store links between similar constructions. Hence, the BrE network of preposition placement constructions must include horizontal relations between stranding and pied-piping constructions.

The present study thus assumes that the stranding and pied-piping constructions described in 3.1.1.1 and 3.1.1.2 form a family of preposition placement constructions which consists of a subfamily of stranding and a subfamily of pied-piping constructions. This constructional family, which is illustrated in figure 3.5, captures the fact that speakers recognize formal similarities between individual stranding constructions, between individual pied-piping constructions, and between stranding and pied-piping constructions. Importantly, the arrow connecting the stranding and the pied-piping families in figure 3.5 does not imply that all constructions included in the pied-piping family also have a corresponding stranded construction in the stranding family, and vice versa. Instead, the arrow indicates that all preposition placement constructions exist at the same hierarchical level, and that speakers store horizontal links between all these constructions. The only exception is, once again, the general stranding construction,

which can be considered slightly more abstract than the other construction types (see 3.1.1.2).

Furthermore, figure 3.5 includes horizontal relations between stranding and pied-piping constructions and their sentence neighbors. For instance, *it*-cleft and topicalization constructions are linked to other members of the information structure family, while the stranded and pied-piped relative clause constructions are connected to the relative family, which includes e.g. subject or object relative clause constructions. The thickness of the lines again reflects the degree of association between constructions. *It*-cleft constructions, for example, which are only superficially similar to relative clauses, are only loosely connected to the relative family, whereas e.g. stranded *that*- or zero-relatives have a much stronger connection to other relatives. An important type of relative clause construction that should be noted in this context are temporal (3.57b) and locative (3.58b) relative constructions in which a “relative adverb” (Biber et al. 1999: 608), such as *when* or *where*, replaces the WH-pronoun and the preposition (Biber et al. 1999: 624; Hoffmann 2011: 37). As shown in (3.57) and (3.58), these constructions can be used as an alternative to stranded and pied-piped *that*-, zero- and WH-relatives (Mußemann 2024: 4).

- (3.57) (a) *All three of them swayed there, lost in the moment **at which** Art chose life* <ICE-NZ:W2F-006#130:1>
- (b) *There is usually a moment **when** her expression switches from sympathetic consideration to one of decisiveness* <ICE-GB:W2F-019 #31:1>
- (3.58) (a) *could you imagine what life would have been like in a place **in which** forty-nine percent of the people didn't want to be there* <ICE-CAN:S2A-034#137:2:A>
- (b) *And this is all so that the finance department can move into the place **where** we already are* <ICE-GB:S1A-082 #62:1:A>

Also note that figure 3.5 is not exhaustive: For example, the stranded WH-interrogative construction is, of course, also related to other interrogative constructions such as subject or object interrogatives. Especially relevant for the present study are adverbial interrogatives such as (3.59) or (3.60) in which, just like in the adverbial relative clauses in (3.57) and (3.58), the preposition and the WH-pronoun are replaced by *when* (3.59b) or *where* (3.60b) (Hoffmann 2011: 47). However, for reasons of space, figure 3.5 does not include all potential relations between preposition placement constructions and their sentence neighbors.

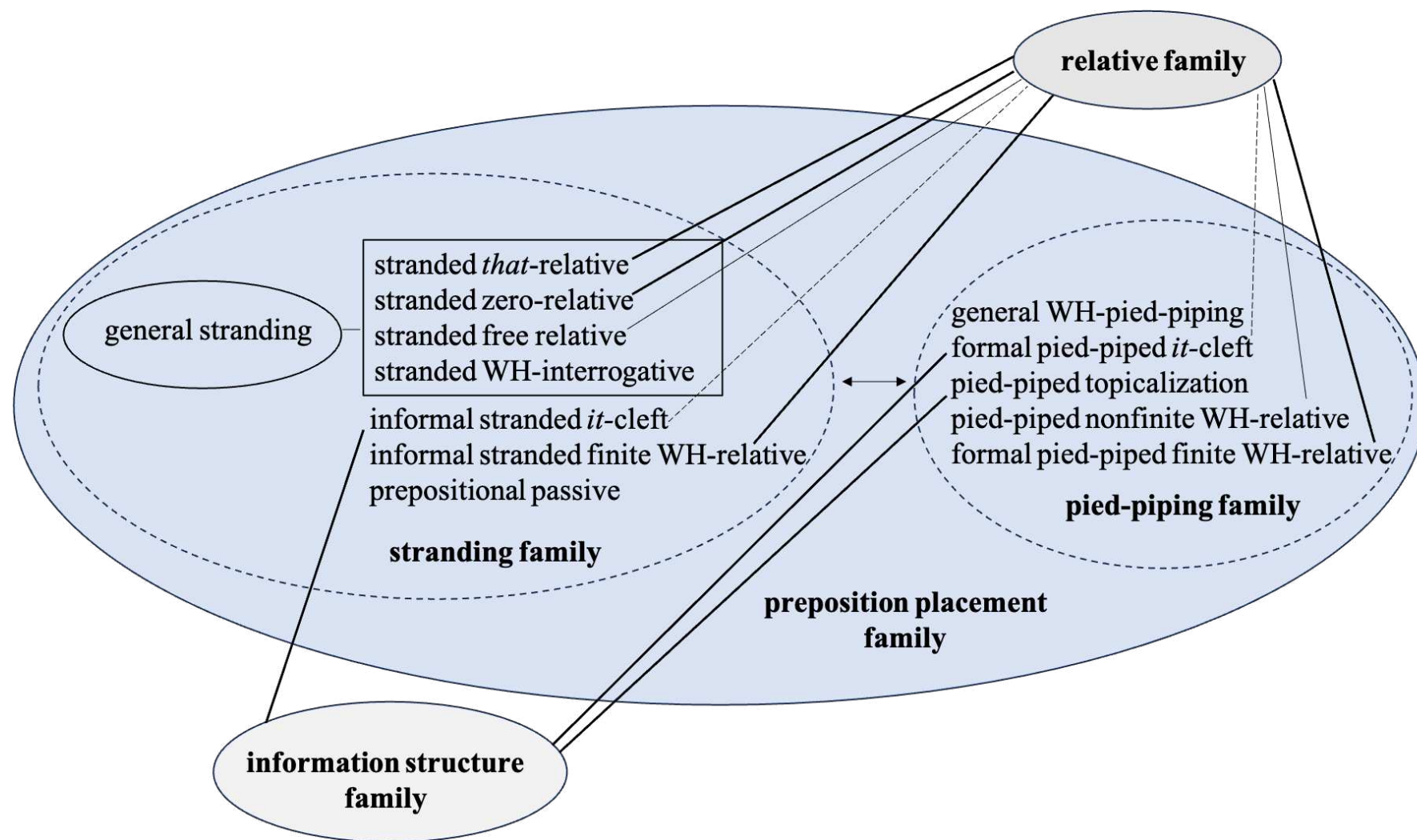


Figure 3.5: The preposition placement family in BrE

- (3.59) (a) *can I ask you **at what stage** did you discover that the Labour ministers were absenting themselves* <ICE-IRE:S1B-056\$A>
- (b) *Michael **when** does the rot set in if there is a rot* <ICE-GB:S1B-024 #83:1:D>
- (3.60) (a) ***on what part of your body** did he hit you with that hammer* <ICE-NZ:S1B-063#118:1:A>
- (b) *And I just thought well now **where** shall I poke him to wake him up* <ICE-GB:S1A-018 #266:1:A>

Finally, it is important to keep in mind that, depending on individual language experience, individual speakers may, of course, store different constructions and different horizontal links between constructions (Schmid 2020: 217). Figure 3.5 should thus be considered a generalized version of the BrE construction network that shows which abstract preposition placement constructions are likely to be entrenched in speakers' mental grammars based on the results of previous research.

3.1.2 The meso-constructional level

In addition to the macro-constructions mentioned in the previous section (3.1.1), speakers also store stranded and pied-piped constructions at a lower level of abstraction, i.e. meso-constructions that consist of both substantive and schematic slots. With respect to preposition placement, research on partly schematic constructions has so far mostly focused on relative clauses (Hoffmann 2011: 160–165; Jach 2021). That both Hoffmann (2011) and Jach (2021) found that speakers store a number of partly schematic pied-piped finite WH-relative clause constructions can be attributed to the fact that WH-relatives strongly favor pied-piping (cf. 3.2.1), which means that speakers are exposed to a lot of positive input for this construction.

Relative clauses are traditionally analyzed as a combination of two independent elements, viz. the antecedent noun head (e.g. *way* in (3.61)) and the relative clause (e.g. *in which the Government behaves*). However, Hoffmann's (2011: 160–165) corpus study showed that BrE speakers store meso-constructional chunks, such as *way in which* (3.61), *frequency with which* (3.62) or *extent to which* (3.63), in which the ANTECEDENT + P + WH string is fully specified.

- (3.61) *But it does actually influence the **way in which** the Government behaves* <ICE- GB:S1B-029 #40:1:A>

- (3.62) *Well the very **frequency with which** John Major's social and educational background have been commented upon is surely proof that we are still obsessed with class* <ICE-GB:S2B-036 #40:1:A>
- (3.63) *The **extent to which** their populations were new and transitory is apparent from the Saint-Lazare terminus's passenger statistics for 1869* <ICE-GB:W2B-002 #68:2>

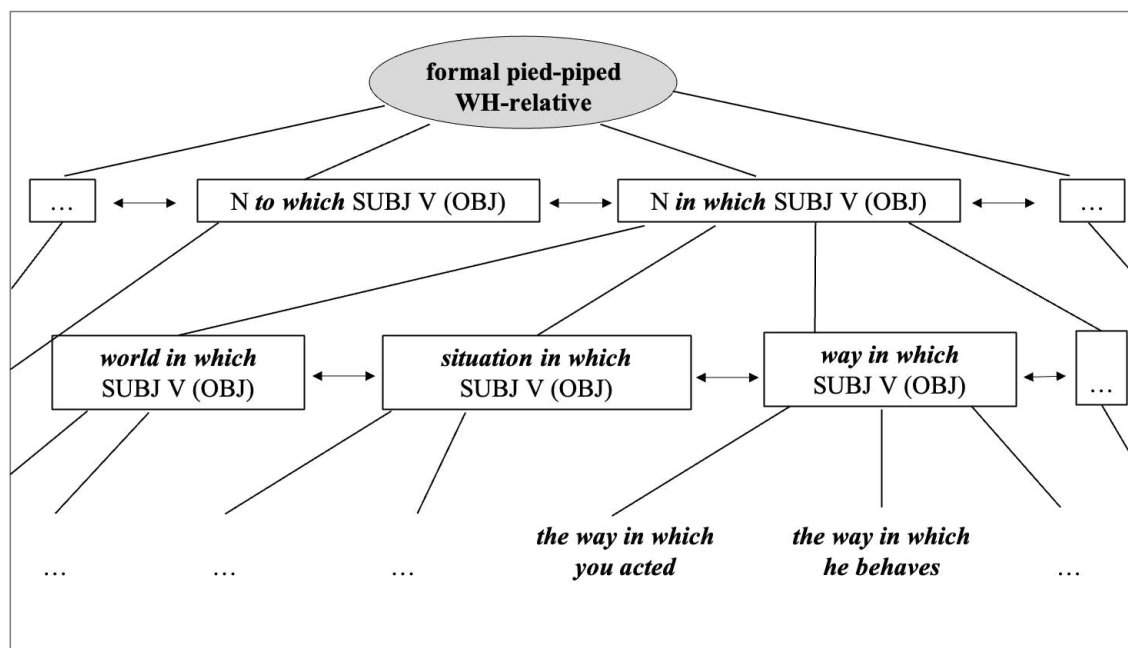


Figure 3.6: Partial construction network of a formal pied-piped finite WH-relative clause construction

This indicates that speakers possess constructions that cut across traditional syntactic categories. As illustrated in figure 3.6, which shows how mental representations of a formal pied-piped finite WH-relative clause construction emerge in a bottom-up fashion, for example, a *way-in-which* meso-construction can become entrenched if speakers are exposed to a number of different constructs that include the string *way-in-which*. Although, from a usage-based perspective, in theory, all ANTECEDENT + P + WH combinations can become entrenched, nouns such as *situation*, *point*, *way*, *manner*, *extent* or *degree* are particularly likely to be involved in pied-piped meso-constructions since these nouns are generally assumed to only license pied-piping (Ross 1986: 132; Johansson & Geisler 1998: 74). This can be explained by the fact that these nouns normally appear in respect (e.g. *situation*), manner (e.g. *way*) or degree (e.g. *extent*) adjunct PPs, which are typically uninterpretable with stranded prepositions (cf. 3.2.2).

Now, if speakers are also exposed to several types of an *in-which* construction, i.e. to constructs such as *the world in which we live* or *the situation in which he succeeded*,

this may lead to the entrenchment of a more abstract *in-which* meso-construction, in which only the P + WH string is phonologically fixed. In a similar way, speakers might develop representations of e.g. a *to-which* relative clause construction (e.g. *the party to which she went*) or other meso-constructional templates in which the P + WH combination is fully specified (Jach 2021: 361).

With respect to other clause types, Hoffmann (2011: 145) found that BrE has an entrenched *how long...for* interrogative construction (3.64) and a *WH...be like* construction (3.65) (Hoffmann 2011: 137–138).

(3.64) *Well so **how** long did you live in Portugal **for*** <ICE-GB:S1A-041 #197:1:B>

(3.65) ***What** was it **like** in Scotland* <ICE-GB:S1A-021 #237:2:C>

Two aspects should be noted with respect to the *WH...be like* construction. First, this construction has a non-compositional meaning which could alternatively be expressed by a *how*-construction (e.g. *How was it in Scotland* in (3.65)). Second, constructions with a *be like* element do not license pied-piping in Standard Englishes (3.66) (Quirk et al. 1985: 818, 1052).

(3.66) * ***Like** what was it in Scotland*

Another meso-construction with a non-compositional meaning that might be entrenched in the mental construction network is the *WH...for* construction (3.67a), which has the meaning ‘why’ (e.g. *Why are you spilling out the water* in (3.67a)). In Standard varieties, the *WH...for* construction does not allow for pied-piping either (3.67b) (Quirk et al. 1985: 818, 1052).

(3.67) (a) ***What** are you spilling the water out **for*** <ICE-IRE:S1A-063\$B>

(b) * ***For** what are you spilling the water out?*

In addition, there are several prepositions, such as *during*, *since*, *beyond*, *under* or *due to*, which normally license pied-piping only (3.68 and 3.69) (Quirk et al. 1985: 817; Johansson & Geisler 1998: 75; Trotta 2000: 63; Hoffmann 2011: 73).

(3.68) (a) *an Act of Remembrance **during which** the wreath is laid* <ICE-GB:S2A-020 #116:2:A>

(b) ? *an Act of Remembrance **which** the wreath is laid **during***

(3.69) (a) *a trestle table **under which** crates of beer were stacked* <ICE-GB:W2F-018 #100:1>

- (b) ? *a trestle table **which** crates of beer were stacked **under***

These lexical effects are frequently attributed to the length and the frequency of these prepositions. It is generally assumed that “it is the most common and the short prepositions which can be [stranded]” (Quirk et al. 1985: 664), whereas infrequent, longer and more complex prepositions require pied-piping (Quirk et al. 1985: 664; Jach 2021: 358). However, so far it remains unclear why more frequent prepositions should favor stranding more than infrequent prepositions (Jach 2021: 359). Another aspect that may account for these lexical effects could be that these prepositions are strongly associated with prototypical adjunct PPs such as frequency (e.g. *during, since*), location (e.g. *beyond, under*) or reason (e.g. *due to*) PPs, which strongly prefer pied-piping (Pullum & Huddleston 2002: 631; Hoffmann 2011: 73; cf. 3.2.2). Regardless of the reason behind these lexical effects, the fact that speakers tend to encounter certain prepositions in pied-piping constructions only may also contribute to the entrenchment of pied-piped meso-constructions, such as a *during-which* or *due-to-which* construction, in which the P + WH string is fixed.

Finally, it is important to note that certain fillers are also strongly associated with certain preposition placement variants. However, in contrast to the idiosyncratic lexical effects of prepositions, the preposition placement preferences of individual fillers can primarily be attributed to a correlation between fillers and clause types (Hoffmann 2011: 41). For instance, *wh-ever* words, such as *whatever, whoever* or *whichever*, occur in free relative clauses (3.70) only (Huddleston, Pullum & Peterson 2002: 1072). Pied-piping in free relatives being a very infrequent phenomenon (cf. 3.1.1.2), this explains why these elements are strongly associated with stranding.

- (3.70) (a) ***whatever meeting** that nomination is publicised **at** shall be considered to be an extension of this meeting* <ICE-GB:S1B-078 #186:1:A>

- (b) ? ***at whatever meeting** that nomination is publicised shall be considered to be an extension of this meeting*

In addition, *what* and *how* are prototypical fillers of WH-interrogatives and free relatives but not of WH-relative clauses (Biber et al. 1999: 87, 608). As a result, speakers are much more likely to encounter *what* and *how* with stranded prepositions than in pied-piping constructions.

The only exception that can indeed be attributed to idiosyncratic lexical effects is the behavior of *who* and *whom*. *Whom* is not only considered more formal than *who*

(Payne & Huddleston 2002: 464), but it is normally also the only acceptable variant in pied-piped WH-interrogatives and WH-relatives (3.71a). At the same time, both *whom* (3.72a) and *who* (3.72b) license stranding (Quirk et al. 1985: 370; Payne & Huddleston 2002: 465).

- (3.71) (a) *the man **on whom** Mrs Thatcher relied* <ICE-GB:S2B-009#59:1:E>
 (b) *? the man **on who** Mrs Thatcher relied*
- (3.72) (a) *He couldn't identify **whom** it belonged **to*** <ICE-CAN:S2B-016#34:1:D>
 (b) ***Who** does the throne belong **to*** <ICE-CAN:S2A-029#149:2:A>

The reason for the stranding requirement of *who* goes back to the synthetic case marking system of OE. While *who* is a nominative form, *whom* can be considered the accusative or oblique version of the pronoun. The ungrammaticality of P + *who* therefore results from the fact that a pied-piped preposition usually requires oblique case, which is not compatible with the nominative case of *who* (Payne & Huddleston 2002: 465; Hoffmann 2011: 42).

As this implies that speakers receive a lot of input for pied-piped *whom*-constructions and for stranded constructions involving the fillers *what*, *how* or *wh-ever*, from a usage-based point of view, the BrE construction network may also contain meso-constructions that capture these associations, such as a stranded *what...P* interrogative or a formal pied-piped P + *whom* relative clause construction.

This overview of partly schematic stranding and pied-piping constructions has demonstrated that the present corpus study will not only have to analyze abstract constructional patterns, but that a comprehensive investigation of preposition placement in World Englishes calls for an exploration of stranding and pied-piping phenomena at all levels of abstraction.

3.2 Preposition placement and processing factors

According to Hawkins' (2004: 3) "Performance-Grammar Correspondence Hypothesis" (PGCH), "[g]rammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments". This can, for instance, be seen in cross-linguistic word order preferences that can be explained by processing constraints (Hawkins 2004; Goldberg 2006: 17; also see e.g. the constraints that have been pointed

out with respect to particle placement, the dative alternation and the genitive alternation in 2.4.1.1). Processing factors are also assumed to be highly relevant in the choice of preposition placement constructions (Gries 2002; Hoffmann 2011: 93–98). As outlined in chapter 1, stranded and pied-prepositions occur in filler-gap constructions, i.e. in constructions which are very difficult to process. In these highly complex environments, pied-piping offers a processing advantage over stranding because it complies with Hawkins’ (1999: 277) Avoid Competing Subcategorizers principle and avoids garden paths. Consequently, constructions with pied-piped prepositions should generally be preferred, and, indeed, more languages license preposition pied-piping than stranding (Hawkins 1999: 277; Hoffmann 2011: 56–58). However, previous studies have identified a number of additional processing-related factors that affect the choice between stranding and pied-piping constructions.

3.2.1 *The complexity of relative clause constructions*

Out of all the constructions discussed in section 3.1, relative clause constructions are the most complex ones. Relative clauses do not only require processing of the FGD but also of the co-indexation domain of the antecedent NP and the corresponding relativizer. For instance, when processing (3.73), a speaker does not just have to establish a relation between the filler and the gap, but also between *several areas* and *on which* (Hawkins 2004: 199; Hoffmann 2011: 96).

(3.73) *I anticipate that there may well be [several areas]_i [on which]_i [the College and AUT would agree]_i* <ICE-GB:W1B-024 #51:4>

The high processing cost of this construction type thus explains why finite WH-relative clauses strongly favor pied-piping, i.e. the variant that is generally easier to process (Trotta 2000: 55–57; Pullum & Huddleston 2002: 628–629; Hoffmann 2011: 155–156).

The most complex relative construction is the restrictive finite WH-relative clause, i.e. a relative construction in which “the reference of the head is a member of a class which can be identified only through the modification that has been supplied” (Quirk et al. 1985: 1239). In the non-restrictive example in (3.74a), the relative clause does not need to be processed in order to establish the reference of *Sarah*. In contrast, the information contained in the relative clause is crucial to identify the reference of *the sisters* in the restrictive construction in (3.74b).

(3.74) (a) *I like Sarah who I'm sharing with but am not too keen on some of the others who are here* <ICE-GB:W1B-009 #51:1>

- (b) *The matron and some of **the sisters with whom she worked at the hospital** remarked on a sparkle in her eye not consonant with the war news* <ICE-GB:W2F-005 #5:1>

This distinction between restrictive and non-restrictive relative clauses is relevant for the competition between stranding and pied-piping. As the interpretation of the antecedent noun in restrictive relative constructions such as (3.74b) is dependent on the parsing of the relative clause, these constructions are more difficult to process than non-restrictive ones (Hawkins 2004: 241). As a result, it is not surprising that Hoffmann (2011: 169–170) found that speakers exhibit an even stronger pied-piping preference with restrictive relative clauses than with non-restrictive ones. While *that*- and zero-relatives are almost always restrictive (for exceptions, see Huddleston, Pullum & Peterson 2002: 1052), finite WH-relative clauses, i.e. a construction type included in the present study, can be both restrictive and non-restrictive (Huddleston, Pullum & Peterson 2002: 1059). Hence, restrictiveness will be a factor that must be taken into account in the corpus analysis. Note, however, that the distinction between restrictive and non-restrictive relative clauses should be viewed as “a gradient rather than as a dichotomy between two homogeneous categories” (Quirk et al. 1985: 1257).

3.2.2 Type of PP

While the effects reported so far indicate that pied-piping offers processing advantages over stranding, depending on the type of PP involved, in English there are also cases in which a stranded preposition can be easier to process than the pied-piped variant. In 1.1, it has already been pointed out that preposition stranding is favored if the verb and the preposition are closely associated with each other. Hoffmann (2011: 155) found that stranding is strongly preferred with prepositional verbs such as *rely on* or *deal with*. In (3.75), for example, *stand for* can be assumed to be stored as a chunk. Therefore, the stranded variant (3.75a), which maintains the verb-preposition string, leads to a reduction in processing cost (Hawkins 1999: 260, fn. 15; Pullum & Huddleston 2002: 629; Hoffmann 2011: 59). The same effect also applies to lexically stored multi-word items, such as *make light of*, *get rid of* or *come down on* (3.76) (Ross 1986: 135; Johansson & Geisler 1998: 77; Trotta 2000: 185; Hoffmann 2011: 155).

- (3.75) (a) *well if the people are selected by the er electorates on the policies **which they stand for*** <ICE-NZ:S1B-024#132:1:S>
- (b) *The full term **for which** an abbreviation **stands** should precede its first use in the text* <ICE-NZ:W2D-007#92:2>

(3.76) (a) *Which side of the argument will we come down on* <ICE-GB:S1B-005 #15:1:A>

(b) *On which side of the argument will we come down?*

Obligatory complements and optional complements have also been found to favor stranding although their stranding preference is not as strong as that of prepositional verbs (Hoffmann 2011: 155). While prepositional verbs such as *stand for* or *deal with* select a specific verb and typically have a non-compositional meaning (Huddleston 2002b: 220; Hoffmann 2011: 68), there are also verbs such as *live* which cannot be used without a complement (* *she lives*) but which allow for different kinds of PPs (3.77a-c). Hence, these PPs can be described as obligatory complements. PPs acting as complements of *be* (3.78) can be considered a special type of obligatory complement since they have been found to have an almost categorical stranding preference (3.78a) (Bailey 1986: 165; Johansson & Geisler 1998: 77; Hoffmann 2011: 138–140). Note that this effect cannot be explained by the semantic weakness of *be* (Trotta 2000: 62) because structures without prepositions (e.g. *I don't know what it is*) license the use of clause-final *be* (Hoffmann 2011: 139). Instead, Hoffmann (2011: 139) attributed the strong stranding preference of *be*-constructions in BrE to “a lexically stored constraint [...] that requires *be* to co-occur with stranded prepositions”.

(3.77) (a) *that's the place **which** he **lives in*** <ICE-GB:S1A-071 #101:1:D>

(b) *And **who** does he **live with*** <ICE-IRE:S1A-049\$B>

(c) *Dieter Schoenfeld is a charter volunteer member of the Loon Survey who with his wife, Marlies, monitors loon populations and reproduction on 10 lakes in the Sudbury area, including McFarlane, **on which** they **live**.* <ICE-CAN:W2B-026#124:1>

(3.78) (a) *that's the box **which** the winch **is in*** <ICE-NZ:S1A-052#163:1:B>

(b) *? that's the box **in which** the winch **is***

In contrast to the obligatory complements in (3.77) and (3.78), the PPs in (3.79) and (3.80) can be considered optional complements because “[e]ven though the omission of the PPs [...] does not result in an unsystematic change in the verb meaning [...], the main verbs can nevertheless be said to subcategorize for the given prepositions” (Hoffmann 2011: 69). For instance, in (3.79) *talk to* cannot be replaced with *talk at* because the *to*-PP “does not simply denote the goal of the event, but refers to [the person] as also taking an active role in the communication” (Hoffmann 2011: 69). In (3.80), the meaning of *write* would not change if the PP was constructed with a different preposition.

Nevertheless, “[t]he choice of the preposition is not arbitrary” (Huddleston 2002b: 228) since verbs such as *speak*, *talk* or *write* “usually realize their theme as an *about*-PP” (Hoffmann 2011: 69).

- (3.79) (a) *I can't remember **who** he was **talking to*** <ICE-GB:S1B-016
#166:1:A>
- (b) *you probably know **to whom** you're **talking*** <ICE-GB:S2A-028
#109:2:A>
- (3.80) (a) *if they're good writers they're good writers no matter **what** they're
writing about* <ICE-CAN:S1A-031#83:1:B>
- (b) *One means in effect the social contract **about which** John Locke
wrote at the end of the seventeenth century* <ICE-GB:S2B-047
#61:1:A>

Now, from a constructionist perspective, verbs which require a specific preposition (e.g. prepositional verbs such as *rely on*), verbs which must be followed by some kind of complement (e.g. *she lives in Africa/with her brother/on a boat*), and verbs which only license specific theme-prepositions (e.g. *talk to*, *write about*) must specify the information about the prepositions or PP types they can be combined with in their lexicon entry (Hoffmann 2022: 158–160; also see Hoffmann 2007b). As this shows that in all of these cases the preposition is syntactically attached to the verb, it can be concluded that preposition stranding is preferred with V + P combinations that are entrenched as lexical chunks. Similarly, Jach (2021: 361) suggested that prototypical stranding constructions involve “specific high-frequency lexical strings such as *go to*, *talk about*, and *aware of*”.

At the same time, there is a group of adjunct PPs, namely respect (3.81), manner (3.82) and degree PPs (3.83), which almost exclusively occur in pied-piped constructions. According to Hoffmann (2011: 141), the almost categorical pied-piping preference of these PP types can be explained by the fact “that stranding a preposition with these leads to an uninterpretable V-P_{stranded} complex [since these PPs] do not add thematic participants to a predicate”. Instead, manner and degree PPs compare events to other potential events (Hoffmann 2011: 141; also see Ernst 2002: 59), while respect PPs establish a reference point in respect of which a clause must be interpreted (Hoffmann 2011: 141; also see Quirk et al. 1985: 483). Hence, the stranded variants in (3.81) – (3.83) are strongly avoided.

- (3.81) (a) *well this is er the situations **in which** the er pakistanis perform
best isn't it* <ICE-NZ:S2A-012#86:1:B>
- (b) *? the situations **which** the Pakistanis perform best **in***

- (3.82) (a) *Again I'm not really concerned so much with meaning but the ways **in which** the satire is <,> achieved* <ICE-GB:S1B-014 #5:1:A>
 (b) *? the ways **which** the satire is achieved **in***
- (3.83) (a) *There are five speed settings for the rate **at which** the display is scanned* <ICE-GB:W2B-039 #52:1>
 (b) *? the rate **which** the display is scanned **at***

Prototypical adjunct PPs such as time (3.84) or location (3.85) PPs also have a strong pied-piping preference. This can be explained by the fact that the interpretation of adjuncts such as *at the time* (3.84) or *in the regions/countries* (3.85), which “relate to the sentence as a whole rather than solely or predominantly to the V and post-V elements” (Quirk et al. 1985: 511), is usually not affected by the verb (Hoffmann 2011: 69). According to Ernst (2002: 328), for instance, time adjuncts “locat[e] an event as a whole in time without making any reference to the event’s internal structure”. Hence, stranded constructions with prototypical adjunct PPs (3.84b, 3.85b), in which the risk of garden path effects is higher than in the corresponding pied-piped variants (3.84a, 3.85a), tend to be avoided due to processing reasons. Nevertheless, the pied-piping preference of time and location PP types is weaker than that of respect, manner and degree PPs (Hoffmann 2011: 155).

- (3.84) (a) *And **at what time** of the day would were uh the holiday-makers allowed to use the swimming pool* <ICE-GB:S1B-066 #150:2:A>
 (b) ***What time** did we start our meeting **at*** <ICE-CAN:S1B-079#305:1:B>
- (3.85) (a) *In 1987 the countries **in which** part-time working was less common were: Italy, Ireland, Luxembourg and Belgium* <ICE-IRE:W2A-019\$A>
 (b) ***what regions** are people homeless **in*** <ICE-IRE:S2A-031\$B>

Adopting a frame semantic approach, Jach (2021: 364) related the different preposition placement preferences of PP complements and adjuncts to temporal iconicity. According to Langacker (2008: 79), temporal iconicity can be described as “a natural tendency for conceived time and processing time to be coaligned, such that the order in which events are conceived as occurring dovetails with the order in which they are conceptualized and described” (Langacker 2008: 79; cited in Jach 2021: 364). Applying this concept to preposition placement, Jach (2021: 364) argued that in stranded constructions, in which the preposition appears closer to the lexical head of the

construction, the placement of the preposition construes the filler “as a focal participant of the [...] event”. In contrast, in pied-piped constructions, the distance between the preposition and the lexical head of the construction suggests that the filler is not a core component of the frame evoked by the verb but backgrounded.

However, as Hoffmann (2011: 65–72, 155) illustrated, a strict two-way distinction between PP complements and adjuncts (as assumed e.g. by Hornstein & Weinberg 1981) fails to take into account a number of PP types with more moderate preposition placement tendencies (also see Johansson & Geisler 1998: 70). For instance, Trotta (2000: 183) claimed that there is “a mixed bag” of adjunct PPs which can be found in both stranding and pied-piping constructions, namely “PPs which have an affected preposition complement, locative PPs expressing goal or target, PPs of accompaniment, and instrumental PPs”. In line with Trotta’s predictions, Hoffmann’s (2011: 157) corpus study found that accompaniment PPs (3.86) and movement PPs (3.87) indeed mildly prefer stranding over pied-piping, which makes them comparable to optional complement PPs.

- (3.86) (a) *then i'm gonna ask them to pass me on to other friends of theirs **who** they work **with*** <ICE-NZ:S2A-040#24:1:A>
 (b) *De Valera was somebody **with whom** they might soon have to work in government* <ICE-IRE:W2A-010\$A>
- (3.87) (a) *we also want to know <,> **which building** we are being moved to* <ICE-CAN:S1B-072#105:1:B>
 (b) *Again from Saint Apollinaro Nuovo this is the figure **towards which** that procession is leading Christ flanked by angels* <ICE-GB:S2A-060 #119:1:A>

According to Hoffmann (2011: 70), this might have to do with the fact that, in contrast to the more prototypical time and location adjuncts, in movement and accompaniment PPs, “the entire PP establishes a thematic relationship with the predicate”. This can be seen in the so-called “*this happened*-test” (see Brown & Miller 1991: 90), which assumes that predicate-independent PPs can be successfully attached to the paraphrase *This happened*. Since the test yields ungrammatical results for accompaniment (3.88) and movement PPs (3.89), these adjuncts can be said to resemble complements in that they do not merely modify events (Hoffmann 2011: 70, 2022: 162).

- (3.88) (a) *John played soccer with Sarah.*
 (b) *John played soccer. *This happened with Sarah.*
- (3.89) (a) *John walked to the car.*
 (b) *John walked. * This happened to the car.*

Moreover, Hoffmann (2011: 157) pointed out that

[f]or movement PPs [the stranding preference] might be explained by the fact that they co-occur with a limited number of motion verbs only (*run, move, walk*, etc.) whose meaning necessarily entails a start and endpoint which is usually encoded by a PP (cf. e.g. *he ran from the church to the school*). Accompaniment PPs, on the other hand, appear with a wide range of predicates (cf. *Bonny worked with Clyde, Angelina was in Africa with Brad, Jude lay on the couch with the nanny*). Yet, due to the fact that they are prototypically only constructed with the preposition *with* and normally introduce animate participants only, a stranded preposition seems easy to parse with these PPs.

Considering that any V + P combination, regardless of the type of PP, has the potential to become entrenched in the mental construction network (Hoffmann 2022: 163), particularly constructions with accompaniment adjuncts may frequently involve lexically entrenched chunks such as *play with* or *work with*, which are easier to process if the preposition is stranded.

At the same time, Hoffmann (2011: 157) found that affected location (3.90) and subcategorized PPs (3.91) exhibit more adjunct-like pied-piping preferences. Unlike prepositional verbs, which select a specific preposition, verbs that combine with subcategorized PPs, such as *keep* or *put* in (3.91), only “require the PP to qualify as a goal” (Hoffmann 2011: 68). Affected location PPs are “limited to a small number of verbs (e.g. *sit, sleep, stand* and *walk*) which clearly affect the interpretation of the prepositional complements, turning them from locations into affected objects” (Hoffmann 2011: 72).

- (3.90) (a) *that too seems to have been plucked uhm almost straight from the sixth-century church even to the sort of uh squashy cushion uh **on which** Christ is sitting* <ICE-GB:S2A-060 #120:1:A>
- (b) ***What're** you gonna cut **on*** <ICE-CAN:S1A-013#79:1:B>
- (3.91) (a) *the building is widened again for the Trinity Chapel and for the so-called Corona, **in which** part of the crown or skull of St Thomas was kept* <ICE-GB:W2B-003 #91:1>
- (b) *that's **what** they were thinking of putting Charlie **on*** <ICE-CAN:S1A-090#220:1:A>

Although constructions with subcategorized PPs and affected location PPs may thus also be likely to contain lexicalized verb-preposition strings (e.g. *put on* (3.91b), *sit on* (3.90a)), the adjunct-like locational meaning which is typical of these structures apparently leads to a relatively strong pied-piping preference for these PP types (Hoffmann 2011: 157).

Figure 3.7 summarizes the preposition placement preferences of all PP types investigated in Hoffmann's (2011: 155) corpus study. It clearly underlines that the widespread assumption that pied-piping is easier to process than stranding typically

applies to more adjunct-like PPs. At the same time, stranding can offer a processing advantage if the verb and the preposition or the PP are closely associated with each other. In order to account for this effect, the corpus data in the present study will be coded according to Hoffmann's (2011) fine-grained classification of PP types.

Stranding	PP type	Examples
	VXP	dinner on Friday night which I'm really looking forward to <ICE-GB:W1B-013 #89:2>
	prepositional X	the policies which they stand for <ICE-NZ:S1B-024#132:1:S>
	optional complement	India whose ambassador I talked to <ICE-GB:S2B-010 #49:1:B>
	movement	places where I went to <ICE-IRE:S1A-100\$B>
	accompaniment	the people who I work with <ICE-CAN:W1B-020#83:5>
	obligatory complement	the place which he lives in <ICE-GB:S1A-071 #101:1:D>
	subcategorized PP	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	the paper upon which they are writing <ICE-GB:W2D-006 #40:1>
	direction	the direction from which the tram would approach <ICE-CAN:W2F-012#53:1>
	location	the room in which they are working <ICE-GB:W2D-006 #93:1>
	time	the moment at which Art chose life <ICE-NZ:W2F-006#130:1>
	frequency	the time for which each part is flooded <ICE-IRE:W2B-030\$C>
	means/instrument	the tools with which they were supposed to be working <ICE-NZ:W2B-009#86:1>
	cause/reason/result	The reasons for which we neglect taking our medicines <ICE-GB:S2B-038 #72:2:A>
	respect	the circumstances in which that initiative came to an end <ICE-NZ:S2B-041#127:1:P>
	manner	the way in which they developed <ICE-GB:S1B-016 #160:1:C>
	degree	the extent to which English displaced Latin <ICE-CAN:W2A-008#51:1>
Pied-piping		

Figure 3.7: Preposition placement preference of individual PP types (based on Hoffmann 2011: 155, 172)

3.2.3 Type of phrase

Another processing-related factor concerns the type of phrase in which the PP is embedded. It is generally assumed that PPs contained in NPs exhibit a stronger pied-piping preference than PPs embedded in verb phrases (VPs) or adjective phrases (AdjPs) because the former are extremely hard to process (Cowart 1997: 16–18; Trotta 2000: 60–63; 184–185; Hoffmann 2011: 84–93; Jach 2021: 354). According to Hawkins’ (2004: 209) principle of “Valency Completeness”, “[t]he human processor prefers FGDs to include the subcategorizer for all phrases within the domain that contain the gap”. Consequently, structures with VP-embedded PPs, in which the human processor can match the filler with the gap upon parsing the main verb, are generally easier to process than constructions in which the gap is contained in an AdjP (e.g. *aware* in 3.92a) or in an NP (e.g. *a copy* in 3.93a). Stranded constructions with AdjP- or NP-embedded PPs are particularly complex because the integration of the filler is delayed even further. In (3.92b), the speaker can only establish a structural relation between the filler and the gap after parsing the preposition *of*, which is contained in the AdjP *aware*. In (3.93b), the filler can only be integrated after the human processor has encountered the preposition *for*, which itself is embedded in the NP *the basis*.

- (3.92) (a) *such statements are not inconsistent with the information [of which]_i we are [aware_i]_{AdjP}* <ICE-IRE:W2D-008\$A>
- (b) *The latter is something [which]_i many healthy people are [aware [of_i]_{PP}]_{AdjP}* <ICE-GB:W2B-023 #64:1>
- (3.93) (a) *The communications to the member states [of which]_i you have received [a copy_i]_{NP}* <ICE-IRE:S2A-033\$A>
- (b) *something [which]_i I think we think the book has laid [the basis [for_i]_{PP}]_{NP}* <ICE-NZ:S2A-023#48:1:C>

While this suggests that stranding should be avoided with both AdjP- and NP-contained PPs, Hoffmann’s (2011: 155) multivariate statistical analysis showed that the pied-piping preference is much stronger for NP-contained PPs than for AdjP-contained PPs, which, surprisingly, turned out to pattern with VP-embedded PPs. One reason for this effect could be that constructions with AdjP-contained PPs frequently consist of copular verbs such as *be* (cf. example (3.92)), which are semantically weak and thus easy to process (Hoffmann 2011: 88; also see Trotta 2000: 62). Furthermore, AdjP-contained PPs tend to involve adjective-preposition combinations that are likely to be lexically entrenched (e.g. *aware of* (3.92), *familiar with*, *responsible for*). Just like prepositional

verbs, these adjective-preposition strings can be expected to favor stranded constructions in which the Adj + P chunk is not broken up (Hoffmann 2011: 88).

Another factor that may contribute to the strong pied-piping probability of NP-contained PPs is a special type of pied-piping in relative clauses. In WH-relative clause constructions in which the PP is contained in an NP which itself belongs to a complex NP, not only the preposition but also one or more NPs in which the PP is embedded can appear in front of the WH-word (Ross 1986: 121; Hoffmann 2011: 89). In (3.94), for example, both the preposition *of* and *the scars*, i.e. the NP in which the PP is contained, are placed in front of *which*.

- (3.94) *I am in hospital awaiting or perhaps recovering from an operation on my neck, [[the scars of which]_i I yet bear_i]* <ICE-CAN:W2F-015#86:2>

Now, in parts, such constructions may be used for semantic reasons since NPs that are pied-piped (e.g. *the scars* in (3.94)) are topicalized. At the same time, this type of pied-piping also leads to a reduction in processing effort because the size of the FGD decreases when additional NPs are moved to the front. Note though that this processing advantage is somewhat reduced by the fact that the size of the co-indexation domain of the antecedent NP and the relativizer automatically increases if additional material appears between the antecedent noun and the WH-word (Hoffmann 2011: 90). Nevertheless, pied-piping constructions such as (3.94) may also explain why NP-contained PPs favor pied-piping more than VP- and AdjP-contained PPs.

In this context, it is also important to mention that relative clauses in which the WH-word is extracted out of subject NPs only license pied-piping (Huddleston, Pullum & Peterson 2002: 1093). While the pied-piped examples in (3.95a) and (3.95b) are grammatical, the stranded version (3.95c) is not an option.

- (3.95) (a) *This payment covers between fifty and sixty per cent of the cost of sending these parcels most of which are food items* <ICE-CAN:S1B-060#26:1:B>
 (b) *these parcels [of which]_i [most_i] are food items*
 (c) * *these parcels [which]_i [most of_i] are food items*

Adopting a processing complexity perspective, Hoffmann (2011: 93) attributed this effect again to Hawkins' (2004: 209) principle of Valency Completeness:

[I]f the human processor prefers to integrate fillers upon encountering the main subcategorizer of a clause, then structures [...] in which a filler must be identified with a gap even before the main subcategorizer has been parsed clearly lead to great processing difficulties and should thus be strongly dispreferred cross-linguistically.

Consequently, in the present study, which focuses on contexts that license variable preposition placement, structures in which the WH-word is extracted out of subjects will have to be excluded. Furthermore, it will be important to include the variable phrase type to account for the fact that NP-contained PPs tend to favor pied-piping more than PPs embedded in VPs or AdjPs.

3.2.4 Complexity of the FGD

Another processing-related factor that affects the choice between stranding and pied-piping has to do with the complexity of the bridging structure, i.e. the material standing between the filler and the gap. It is commonly assumed that constructions with long FGDs, which require that the filler has to be held active in working memory over long distances, are more difficult to process than shorter ones (Gibson 1998: 8, 11; Hawkins 1999: 251). In addition, the human processor prefers FGDs that reduce both syntactic and semantic processing (Hawkins 2004: 197–203). For instance, cross-linguistically, speakers seem to prefer semantically weak and unspecific bridging verbs such as *think* or *say* over more specific and semantically more complex verbs such as *whisper* or *realize* (e.g. *What did John **say** that Mary talked about?* is preferred over *What did John **whisper** that Mary talked about?*; Hawkins 2004: 201). In a similar vein, Kluender (1990: 199) claimed that open-class elements that appear in the path from the filler to the gap require more processing cost than closed-class elements.

Since pied-piping is generally easier to process than stranding (cf. 1.1), long and semantically complex bridging structures, in which additional material has to be parsed before the filler can be matched with the gap, should favor pied-piping (Johansson & Geisler 1998: 76; Hawkins 1999: 251, 2004: 201). Support for this assumption comes from Gries' (2002: 237–238) corpus study, which identified the barrierhood and the length of the bridging structure as significant factors that lead to an increase in pied-piping. That complex bridging structures involve an increase in processing cost was also confirmed by Hoffmann's (2011: 193) magnitude estimation experiments, in which BrE speakers rated stranding in simple relative clauses such as (3.96a) better than in more complex ones such as (3.96b) and (3.96c), in which the relative clause *whom Jane relied on* is embedded in another clause. This suggests that constructions with syntactically complex bridging structures might be used rarely in natural language use (examples in (3.96) from Hoffmann 2011: 193).

- (3.96) (a) *I saw the teacher **whom Jane relied on**.*

- (b) *I saw the teacher [whom you claimed] Jane relied on.*
- (c) *I saw the teacher [whom you claimed that] Jane relied on.*

At the same time, Hoffmann’s experiments demonstrated an important interaction effect between complexity and PP type: Even with complex constructions such as (3.96b) and (3.96c), speakers rated stranding better than pied-piping, which can be explained by the fact that all of Hoffmann’s stimuli items included lexicalized V + P combinations such as *rely on*. Since these verb-preposition strings are broken up in pied-piped constructions, stranding is favored even with syntactically complex bridging structures (Hoffmann 2011: 202). Similarly, Hoffmann’s (2011: 168, 224–225) corpus study indicated that pied-piping is only preferred in long filler-gap constructions with more adjunct-like PPs, whereas long constructions with prepositional verbs favor stranding (also see Trotta 2000: 188). This interaction of complexity and PP type will have to be kept in mind in the present study.

3.2.5 Other processing-related factors

This section will focus on other processing-related factors which may have an effect on preposition placement but which have so far received less attention in the literature.

One such factor that should be investigated in the present analysis is voice. The only study that has explored the effect of voice in connection with preposition placement is Gries’ (2002: 237–238) multivariate analysis based on data from the British National Corpus (BNC), which found that preposition stranding is dispreferred when the main verb is in the passive (3.97b). Arguing that the passive is harder to process than the canonical active, Gries (2002: 237–238) interpreted this finding as a processing effect.

- (3.97) (a) *The Family Practitioner Committee (FPC) [to whom]₄
[we]₁ [will send]₂ [your form]₃*
- (b) *The Family Practitioner Committee (FPC) [to whom]₄ [your
form]₃ [will be sent]₂ (by us)₁ <ICE-GB:W2D-001 #23:1>*

Now, on the formal level, the passive construction differs from the active construction in that the verb phrase consists of a *be*-form followed by the past participle of a full verb (e.g. *will be sent* in (3.97b); Hoffmann 2022: 203). Furthermore, actives and passives can be considered “different construals of the same scenes” (Hoffmann 2022: 203). For example, both (3.97a) and (3.97b) describe the same event. However, in the active construction in (3.97a), it is the agent (i.e. *we*) that appears in the subject position because the scene is viewed from “the vantage point” of the agent (Hoffmann 2022: 203).

In contrast, in (3.97b), the agent is not mentioned explicitly at all (as the *by*-PP is an optional element). Instead, the passive construction, which “construe[s] the most affected, patient-like participant as the vantage point of the scene” (Hoffmann 2022: 203), places the affected element (i.e. *your form*) in subject position. This shows that the passive construction can only be used if “there is an event that can be construed as involving an agent doing something that affects a patient” (Hoffmann 2022: 205). Furthermore, passives are typically associated with academic texts, while they hardly appear in informal registers such as conversation (Biber et al. 1999: 476). As this implies that passive constructions surface less frequently in speakers’ input than active constructions, the passive should generally be less deeply entrenched in speakers’ mental grammars. Since the use of the passive can thus be expected to involve more processing cost than the active, passive constructions might favor pied-piping.

Moreover, a wide range of psycholinguistic studies (e.g. Clark & Clark 1977: 105–106; Davison & Lutz 1985; Ferreira 2003; Wanner 2009; Kharkwal & Stromswold 2014; Street & Dąbrowska 2014) as well as research on L1 (e.g. Baldie 1976; Maratsos et al. 1985; Aravind & Koring 2023) and L2 acquisition (e.g. Marinis & Saddy 2013; Mirdamadi & De Jong 2015) have demonstrated that the passive is difficult to process as it is “more complex morphologically, and the expected linking of semantic role and syntactic position (agents are normally subjects, themes are direct objects) is broken up” (Wanner 2009: 14). There is ample evidence that the acquisition of the passive is delayed compared to the active (de Villiers & de Villiers 1973; Baldie 1976; Pinker, Lebeaux & Frost 1987), and that the interpretation of the passive can cause difficulties for children (Maratsos et al. 1985; Aravind & Koring 2023). Street & Dąbrowska (2014) showed that the passive is frequently misinterpreted even by adult L1 speakers (also see Ferreira 2003). Furthermore, the cognitive complexity of the passive is reflected in the fact that speakers’ reaction times are slower for passives than for actives (Davison & Lutz 1985; Berndt et al. 2004; Kharkwal & Stromswold 2014; Street & Dąbrowska 2014), and that the passive is associated with higher hesitation occurrence (Engelhardt et al. 2010; Mirdamadi & De Jong 2015).

While passives are thus widely known to make high processing demands on speakers, the impact of voice on the choice between stranding and pied-piping clearly requires further investigation. As Hoffmann (2011: 94) pointed out, Gries’ (2002) results must be taken with a grain of salt. Although the BNC has a hundred times as many verbs as the ICE-GB, Gries only found 179 stranding tokens in the BNC, while 202 prepositions are tagged as stranded in the ICE-GB. Gries thus seems to have used some kind of

unreliable search algorithm which did not identify all stranded prepositions. According to Hoffmann (2011: 94), “[s]uch an approach, however, seriously affects the validity of his statistical analysis since his results are based on a highly skewed dataset which in all likelihood ignored a great number of relevant stranded tokens”. Moreover, Gries (2002) only included interrogative tokens in his statistical analysis. The present study will thus have to investigate whether voice also emerges as a significant variable when looking at more complex construction types such as relative clauses (cf. 3.2.1).

In addition to the effect of voice, Gries (2002: 234) also analyzed whether the frequency, the concreteness, the agentivity or the degree of entrenchment of the head noun of the extracted phrase have an effect on preposition placement. Since his study aimed to test the hypothesis that the choice between stranding and pied-piping “will [...] be sensitive to the processing cost incurred by the planning and production of the utterance” (Gries 2002: 232), Gries apparently considered all these variable to be processing-related, but provided no further explanations for this assumption. However, a closer look at his variables AGENT-HEAD, CONCRETE-HEAD, FREQ-HEAD and ENTRENCH-HEAD reveals that they all somehow encode the same concept, namely the degree of entrenchment of the head noun of the extracted phrase. Not only are more frequent nouns typically more deeply entrenched than infrequent ones, but concrete words with transparent form-meaning mappings can also be learned more easily than more abstract terms (Schmid 2020: 207–208). Moreover, it is generally assumed that “[l]inguistic elements and structures that are associated with concrete and tangible experience offering rich mental images are more likely to become entrenched easily and quickly than more abstract kinds of experience” (Schmid 2020: 222). More frequent and concrete nouns should thus be more deeply entrenched in speakers’ mental grammars than infrequent, abstract items. The fact that all of Gries’ variables predict a very similar phenomenon could explain why his general linear model exhibited problems of multicollinearity (Gries 2002: 240), and, consequently, also why none of these variables turned out to be significant in his analysis (Gries 2002: 237).

Nevertheless, in a study that focuses on processing effects, the degree of entrenchment of the filler is an interesting variable that deserves further investigation. From a processing perspective, deeply entrenched filler phrases may be more likely to co-occur with stranded prepositions than with pied-piped ones. As deeply entrenched fillers can be activated easily, they may have a higher chance of being used clause-initially, i.e. in stranding constructions, than filler phrases that are less deeply entrenched. However, in the present study, which includes all clause types that allow for variable

preposition placement, it must be kept in mind that Gries’ (2002) variables are only relevant for constructions with complex filler phrases involving a nominal head (e.g. [*Which house*] *do you live in?*) but not for e.g. WH-relative clauses, which typically only consist of a one-word filler (e.g. *the house* [*which*] *I live in*). Instead of the four variables proposed by Gries (2002: 234), the present study will thus simply use the filler frequencies to measure the degree of entrenchment of the filler.

3.2.6 *Summary: An overview of prototypical preposition placement constructions*

Since the present analysis aims to explore if processing factors have a stronger effect on the constructional choices of speakers at lower stages of Schneider’s model, it will investigate if there is a correlation between the preference for prototypical constructions and Dynamic Model phases. In the light of Hawkins’ (2004: 3) PGCH, which claims that “[g]rammars have conventionalized syntactic structures in proportion to their degree of preference in performance”, in the following, the term “prototypical constructions” will be used to refer to linguistic structures that should be preferred from a processing perspective. Table 3.2, which summarizes the main points of the previous sections on the role of processing factors in the competition between stranding and pied-piping, gives an overview of the preposition placement constructions that will be considered prototypical in the present corpus study.

Table 3.2: Prototypical stranding and pied-piping constructions

Prototypical stranding constructions	Prototypical pied-piping constructions
	Pied-piped (restrictive) finite WH-relative clause construction
Stranded constructions with lexicalized verb-preposition strings	Pied-piped constructions with time, location, direction, instrument, contingency, frequency, respect, manner, and degree PPs
	Pied-piped constructions with NP-embedded PPs
	Pied-piped passive constructions ⁴

⁴ Note that the term “pied-piped passive construction” must not be confused with the term “prepositional passive construction” used in 3.1.1.2. While the latter expression is used to refer to constructions with stranded prepositions in which the NP appearing in the subject slot also acts as the prepositional complement of the stranded preposition (e.g. *King’s Canterbury is being spoken of very highly*), the pied-

With respect to finite WH-relative clause constructions, i.e. the most complex construction type that will be analyzed, both restrictive and non-restrictive pied-piping constructions will be considered more prototypical than the stranded variants (cf. 3.2.1). However, since the interpretation of the antecedent head in restrictive relative clauses is dependent on the parsing of the relative clause, restrictive finite WH-relative clause constructions with pied-piped prepositions will be considered even more prototypical than non-restrictive ones.

In addition, table 3.2 clearly shows that there are much more prototypical contexts for pied-piping than for stranding. A stranded preposition is only easier to process than a pied-piped preposition if a construction contains a lexicalized verb-preposition string. If the verb and the preposition are closely associated with each other, stranding should therefore even be favored with very long FGDs (cf. 3.2.4). At the same time, pied-piping should be preferred in all other environments that somehow increase processing complexity, namely in finite WH-relative clause constructions, with prototypical adjunct PPs, with PPs that are themselves embedded in NPs, and when the main verb appears in the passive.

3.3 Preposition placement constructions in World Englishes

As pointed out in chapter 2, input effects, processing constraints, second-language learning principles and cross-linguistic influence can lead to constructional innovation in PCEs. This section will give an overview of previous studies that pointed out ways in which such factors may lead to cross-varietal differences in the use and entrenchment of stranding and pied-piping constructions. While 3.3.1 will outline the main points of Hoffmann's (2011) usage-based comparison of the BrE and the KenE construction networks, in 3.3.2 and 3.3.3 other studies will be sketched that found innovative uses of preposition placement constructions in L2 Englishes.

3.3.1 *The BrE vs. the KenE construction network*

As briefly mentioned in section 1.1, Hoffmann (2011) found that differences between the BrE and KenE networks of preposition placement constructions can be attributed to two main factors: processing constraints and formality effects. That KenE speakers have a

piped passive construction describes structures with a pied-piped preposition in which the main verb is in the passive (e.g. *the committee to whom your form will be sent*).

stronger preference for constructions that are easy to process was shown in Hoffmann's (2011: 182–187) magnitude estimation experiments, in which BrE speakers and KenE speakers were asked to judge the acceptability of stranding constructions with prepositional verbs, with time and location PPs, and with manner and frequency PPs. Figure 3.8 shows that BrE speakers rated preposition stranding with prototypical adjunct PPs such as time or location PPs almost as good as the grammatical fillers. In contrast, in KenE (cf. figure 3.9), in which stranding with time and location PPs was judged as bad as stranding with manner and frequency adjuncts, all stranded constructions with adjunct PPs received significantly lower scores than the grammatical fillers (Hoffmann 2011: 182, 187, 224). For the KenE speakers, stranding thus only seems to be acceptable with lexicalized V + P combinations. In contrast, pied-piping, which avoids garden paths, is strongly preferred in all other contexts (Hoffmann 2011: 187). Hence, Hoffmann (2011: 272) concluded that “in Kenyan English the prototypical associations of stranding with prepositional verbs or pied-piping with adjunct PPs appears to be much stronger than in British English”.

In addition, speakers of BrE and KenE are not affected equally by stylistic variables. This effect is illustrated in figures 3.10 and 3.11, which summarize Hoffmann's (2011: 167–168) usage-based account of the BrE and the KenE construction networks.⁵ While BrE can be said to have a formal pied-piped WH-relative clause construction and an informal stranded finite WH-relative clause construction as well as a formal pied-piped cleft construction and an informal stranded cleft construction, the KenE network only contains a stranded cleft construction and a pied-piped finite WH-relative clause construction. Now, the absence of a pied-piped cleft construction in KenE can be explained by the low token numbers of clefts in ICE-Kenya (Hoffmann 2011: 267). At the same time, the lack of an informal stranded finite WH-relative clause construction can be attributed to the fact that “[p]reposition-stranding in Kenyan English is [...] less productive and limited to fewer contexts than in British English” (Hoffmann 2011: 224). Since the pied-piped finite WH-relative clause construction is easier to process than the stranded variant (cf. 3.2.1), the L2 speakers, who receive less input, seem to strongly prefer the prototypical constructional choice (Hoffmann 2011: 268). Consequently, they use the pied-piped finite WH-relative clause construction regardless of the context. As Hoffmann (2011: 267) pointed out, in KenE, “it is still possible to construct stranded

⁵ In Hoffmann's construction networks, circles mark constructions that are likely to be entrenched from a usage-based perspective, while boxes are used for constructions with idiosyncratic properties (Hoffmann 2011: 266).

finite WH-relative clauses online, but since the pied-piped alternative is stored it will become available more quickly even in less formal contexts”. Besides this, the strong preference for pied-piped finite WH-relative constructions in KenE may also be explained by formal instruction, which may prime the L2 speakers towards pied-piping (Mußemann 2024: 5). For the present analysis, Hoffmann’s (2011) findings thus suggest that differences in the constructional choices between L1 and L2 speakers may result from a complex interaction of processing factors and stylistic effects.

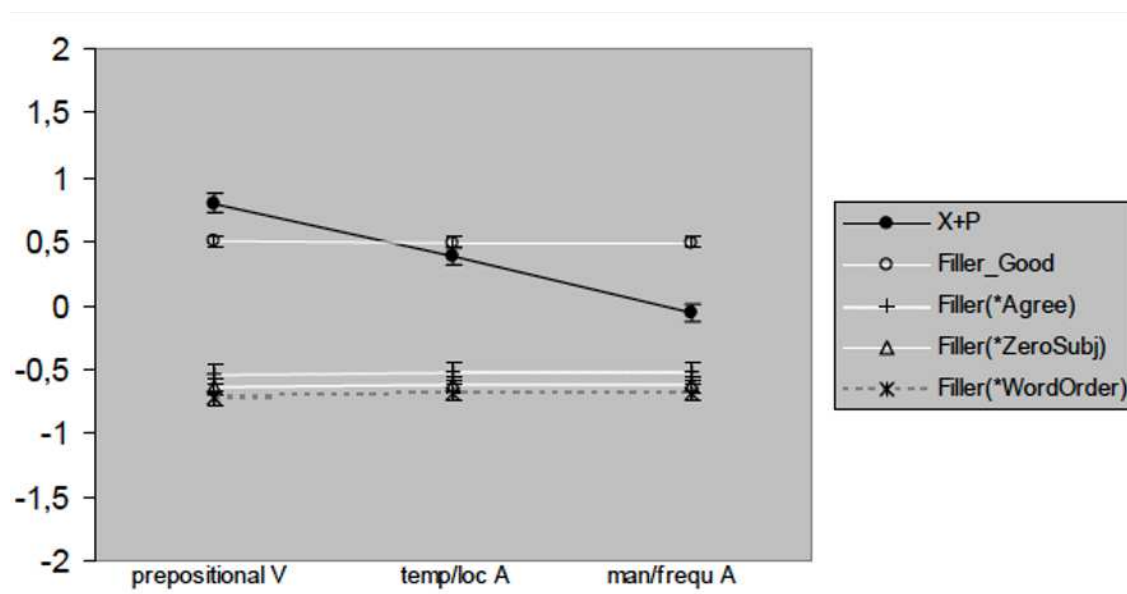


Figure 3.8: Magnitude estimation scores for stranding across different PP types (BrE) (Hoffmann 2011: 182)

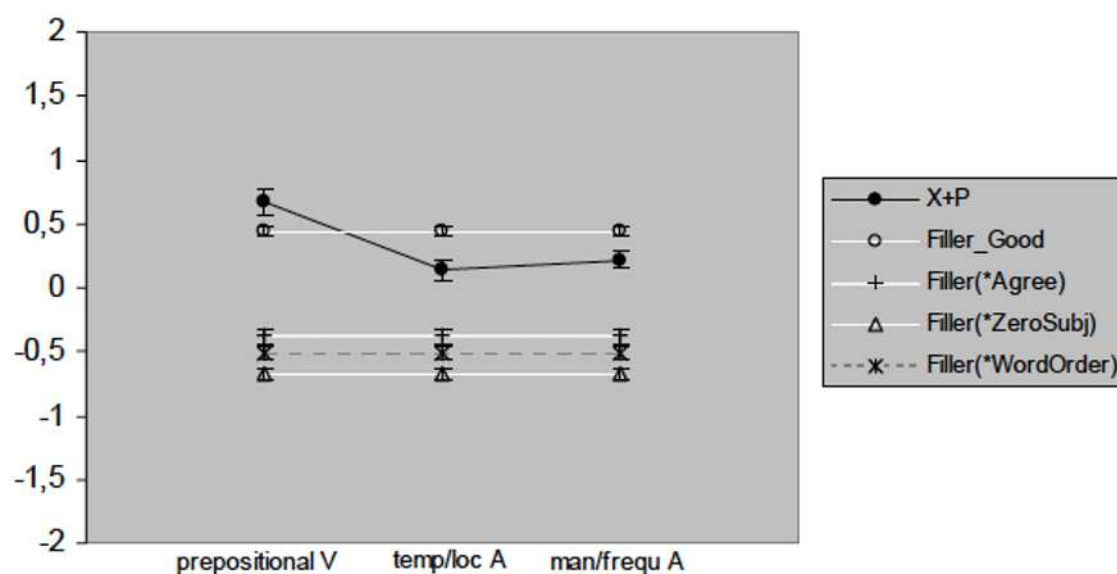


Figure 3.9: Magnitude estimation scores for stranding across different PP types (KenE) (Hoffmann 2011: 188)

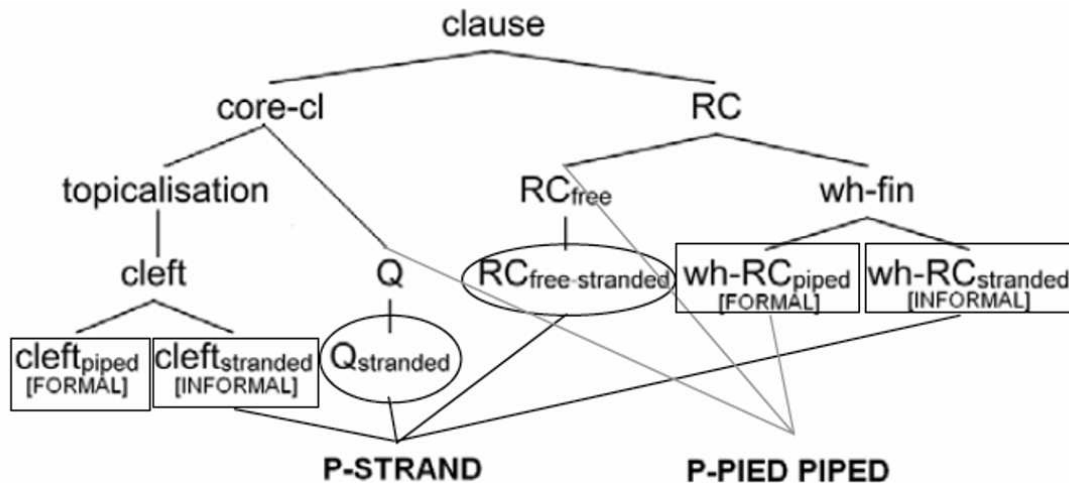


Figure 3.10: Usage-based construction network for BrE (Hoffmann 2011: 267)

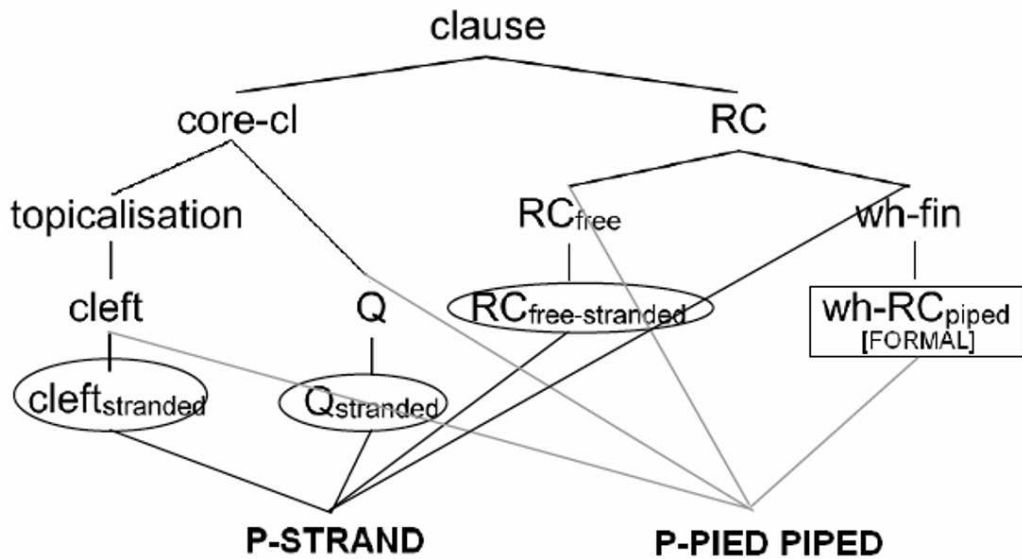


Figure 3.11: Usage-based construction network for KenE (Hoffmann 2011: 268)

Furthermore, the construction networks of the two varieties differ with respect to the entrenchment of meso-constructions. The fact that Hoffmann (2011: 160–165) identified more entrenched ANTECEDENT + P + WH combinations such as *speed with which* or *degree to which* for BrE than for KenE can be attributed to the fact that BrE speakers are exposed to more input. Moreover, in line with the finding that KenE speakers rate stranding with adjunct PPs lower than BrE speakers, Hoffmann’s corpus study (2011: 145) demonstrated that BrE has entrenched the stranded interrogative string *how long...for* (e.g. *How long did you live in Portugal for?*; cf. 3.1.2), while his KenE data only contained the pied-piped variant *for how long*, which is easier to process. In addition, Hoffmann (2011: 140) found that the idiomatic *WH...be like* construction (cf. 3.1.2) is not as deeply entrenched in KenE, and that obligatory complements with *be*, which are

almost categorical stranding environments in BrE (cf. 3.2.2), are less strongly associated with stranding in the L2 variety.

Hoffmann's (2011) study thus indicates that speakers of BrE have more entrenched meso- and macro-constructions than speakers of KenE, who encounter stranding and pied-piping constructions with lower frequencies. In the light of these results, it can be expected that the varieties investigated in the present study will differ from each other with respect to the entrenchment of stranding and pied-piping constructions due to the specific type and amount of input that is associated with different evolutionary stages of Schneider's model.

3.3.2 *Quantitative differences between L1 and L2 Englishes*

Apart from Hoffmann's (2011) analysis of BrE and KenE, there is a range of other studies on preposition placement that found quantitative differences between L1 and L2 Englishes. Following Laporte (2021: 48), quantitative innovations can be defined as features that

[r]efer to differences in frequency of use of conventional variants in comparison with the benchmark variety, which [...] is generally the input variety. They are thus features that are relative in nature, i.e. that do not represent a break with conventions of the source variety. [...] Preferences for some syntactic pattern and relative differences in constraints on the use of syntactic constructions are subsumed under 'quantitative features'.

In the literature reviewed, quantitative innovations in the use of preposition placement constructions are mostly attributed to two factors: input effects and cross-linguistic influence.

That input is a factor that can be expected to lead to cross-varietal differences in the competition between stranding and pied-piping has already been pointed out in connection with Hoffmann's (2011) investigation of KenE (cf. 3.3.1). In addition, a number of studies suggested that stranding is favored by learners of English because of its salience in speakers' input. Bardovi-Harlig (1987) investigated the acquisition of stranded and pied-piped WH-relatives and WH-interrogatives by L2 learners of English. Although stranding is rare cross-linguistically and thus more marked than pied-piping (cf. 2.4.3.1), she found that all learners, regardless of the language background, acquired preposition stranding before pied-piping. Defining salience "in terms of availability of data", Bardovi-Harlig (1987: 401–402) linked this effect to the high salience of preposition stranding. Now, the claim that learners should receive a lot of input for stranded WH-interrogative constructions is not surprising since WH-interrogatives strongly favor stranding for discourse-functional reasons (cf. 3.1.1.2). At the same time, WH-relatives are normally associated with pied-piping due to their complexity (cf. 3.2.1).

However, according to Bardovi-Harlig (1987: 402), speakers also encounter a lot of input for stranded prepositions in relative constructions because of the stranding requirements of *that*- and zero-relatives.

Bardovi-Harlig's (1987) finding that learners acquire stranding before pied-piping was confirmed by several studies (see Kao 2001 on Japanese learners; Sadighi, Parhizgar & Saadat 2004 on Iranian learners; Rezai 2006 on Persian learners; Shin, Yoon & Chung 2016 on Korean learners). Moreover, Jach (2018: 286) found that German and Chinese learners of English rate stranding in WH-relative clauses more acceptable than pied-piping, arguing that

the L2 learners in this study had likely become attuned to the marginal relative frequencies of preposition fronting and stranding, which was reflected in their acceptability ratings. Because the learners had experienced stranding more often than fronting across all clause type contexts in language use, preposition stranding had become more entrenched and also more acceptable to them than fronting. (Jach 2018: 292)

The results of these studies could thus be taken as an indication that the acquisition of stranding is facilitated by the salience of stranding in the input of learners. As speakers are likely to recognize similarities between e.g. stranded *that*- and stranded WH-relative clause constructions (cf. 3.1.1.2), entrenched V + P chunks that are typical of the former structures may also have a high chance of being activated in the latter constructions. However, as pointed out in section 3.1, from a usage-based perspective, it is very implausible that speakers acquire a single abstract macro-construction that is used to produce stranded *that*-, zero-, and WH-relative clause constructions. Instead, it is important to note that the studies of Bardovi-Harlig (1987), Kao (2001), Sadighi, Parhizgar & Saadat (2004), Rezai (2006), Shin, Yoon & Chung (2016), and Jach (2018) only investigated stranding and pied-piping with prepositional verbs. Considering that lexicalized V + P items strongly favor stranding for processing reasons (cf. 3.2.2), the strong stranding preference observed by these studies may also be interpreted as a processing principle (also see Hoffmann 2011: 104).

Furthermore, at least in some cases, cross-linguistic influence seems to lead to a pied-piping preference in learner Englishes. One of the earliest studies that made such a claim was Mazurkewich's (1985) analysis of preposition placement in WH-interrogatives, in which Quebec French speakers, whose L1 only permits pied-piping, turned out to produce more pied-piped prepositions than speakers of Inuktitut, a major Inuit language of Canada which does not have prepositions. In addition, Salehi (2011) conducted a study in which Iranian L2 learners had to complete finite WH-relative clauses with stranded or pied-piped prepositions. The fact that advanced learners were more likely to use pied-piped prepositions than intermediate learners may be seen as an

indication that the learners' constructional choices were influenced by Persian, a language which permits pied-piping only (Salehi 2011: 96). A similar observation was made by Almahammed, Ariff & Sidek (2015), who argued that the frequent use of pied-piping by Jordanian L2 speakers may reflect cross-linguistic influence from Arabic (Almahammed, Ariff & Sidek 2015: 24).

As pointed out above, Jach (2018) attributed the stranding preference of learners to the fact that learners can be expected to receive more input for stranded prepositions than for pied-piping constructions. In addition, both his (2018) magnitude estimation experiments and his (2021) corpus study provided evidence for cross-linguistic influence. In Jach's (2018: 286) acceptability study, German learners, whose L1 licenses pied-piping only, judged pied-piping constructions better than Chinese learners, whose L1 does not have relative constructions with stranded or pied-piped prepositions. Likewise, his (2021: 361) corpus study found that European learners have a stronger preference for pied-piping than Japanese, Korean or Chinese learners, who have L1s which tend to omit prepositions in relative clauses (Jach 2021: 345). Drawing on Höder's (2012, 2014) concept of diaconstructions, Jach (2021: 362) therefore concluded

that the European learners have acquired an interlingual fronting construction which represents the overlap of fronting [prepositional relative clauses] in English and their respective native language and which they use as highly schematic template for placing prepositions in English. As a consequence, the acquisition of English fronting was probably boosted.

While it is doubtful whether speakers indeed acquire such an abstract, language-unspecific pied-piping construction (cf. 2.4.3.2), these findings show that the present analysis will have to take into account the influence of local languages. Speakers who receive a lot of input for pied-piped prepositions in their L1s may also favor pied-piping constructions in English.

3.3.3 *Qualitative innovations*

In addition to the quantitative effects described in the previous section, there is also evidence that L2 speakers make use of preposition placement constructions which differ qualitatively from the stranding and pied-piping constructions that are conventionalized in Standard varieties. Following Laporte (2021: 48), from a constructionist perspective, a qualitative innovation can be described as “a form-meaning mapping that does not conform to the conventional form-meaning mappings of the benchmark variety”.

Hoffmann (2011: 133–134) found four different types of non-standard preposition placement in ICE-Kenya, namely constructions with doubled (3.98), missing (e.g. *aware* instead of *aware of* in (3.99)) and unexpected (e.g. *focus at* instead of *focus on* in (3.100a);

conditions to which instead of *conditions under/in which* in (3.100b)) prepositions as well as structures in which a resumptive pronoun appears after a preposition that would otherwise be stranded (e.g. *into it* in (3.101); examples in (3.98) - (3.101) from Hoffmann 2011: 133).

- (3.98) *The thing is **to whom** this drug is going **to** and the use it is going to be put into* <ICE-EA-K:S1B025K>
- (3.99) *suggesting that most of the drugs that come into this country the authorities are **aware** but a few people are bribed so that they* <ICE-EA-K:S1B004K>
- (3.100) (a) *I think we what you are basically trying to **focus at** is that we are narrowing down to the problem of malsocialisation* <ICE-EA-K:S1B031K>
- (b) *Hostels no longer provide **conditions to which** students can study efficiently* <ICE-EA-K:S2B032K>
- (3.101) *It is a process you have to get initiated into **it*** <ICE-EA-K:S1A026K>

Furthermore, in his magnitude estimation experiments, KenE speakers rated syntactically complex constructions with doubled prepositions such as (3.102) better than BrE speakers (Hoffmann 2011: 210, 225; example from Hoffmann 2011: 189; own emphasis).

- (3.102) *I saw the teacher **on** whom you claimed that Jane relied **on**.*

This clearly reflects a processing principle. In constructions with prepositional verbs (e.g. *rely on* in (3.102)), stranding is preferred from a processing perspective because the in-situ placement of the preposition facilitates the processing of the lexicalized verb-preposition string. At the same time, the use of a pied-piped preposition, which reduces the risk of garden paths effects, also offers a processing advantage (Hoffmann 2011: 204; also see Nykiel 2010: 159). Moreover, the use of doubling structures seems to be linked to the entrenchment of P + WH and V + P chunks. With respect to (3.98), for example, both the strings *to whom* and *go to* are likely to be stored as independent items in the construction network. Since entrenched chunks have a high chance of being activated, doubled preposition structures may frequently result from “a lack of inhibitory control” (Jach 2021: 366).

Entrenched constructional chunks are also highly relevant to account for another type of preposition doubling, which will be referred to as “mismatching” (Riley & Parker 1986: 293) in the present study. In contrast to “doubling” structures, in which the same

preposition appears twice, in constructions with mismatched prepositions such as (3.103), speakers use two different prepositions (e.g. *in* and *to* in (3.103)).

(3.103) *The comparison between them will start with the classification of types of States **in which they belong to*** <ICE-HK:W1A-010#7:1>

Riley & Parker (1986: 302) found that mismatching structures frequently involve the high-frequency preposition *in*. This assumption also applies to example (3.103), in which both the string *in which* and the prepositional verb *belong to* are high-frequency items which should be entrenched in the construction network. Structures with mismatched prepositions can thus be said to result from a blend of two entrenched chunks (e.g. *in which* and *belong to*) which are both “locally acceptable” (Hoffmann 2011: 274) but which lead to an innovative preposition placement structure when being combined.

Doubling and mismatching structures have been found in several studies on preposition placement in L2 Englishes (Bardovi-Harlig 1987: 399; Shin, Yoon & Chung 2017: 103; Jach 2021: 350). It should be noted, though, that instances of preposition doubling such as *the world in which we live in* also occur in L1 Englishes (Riley & Parker 1986; Radford 2004: 192, 2019: 133; Jach 2021: 350). However, as speakers of L2 Englishes can be expected to be more strongly affected by processing constraints and to rely more on entrenched chunks, in the present analysis, these innovations should surface more frequently in PCEs.

Processing factors can also explain the occurrence of resumptive pronouns. Structures with resumptive pronouns such as (3.101) are easier to process than constructions with stranded prepositions because a resumptive pronoun explicitly indicates how the filler and the gap are related (Hoffmann 2011: 106). Particularly in highly complex environments such as long filler-gap constructions, the use of resumptive pronouns may thus be preferred as it leads to a reduction in processing cost. Support for this assumption comes from L1 acquisition. According to Diessel (2004: 121), in early childhood speech,

the occurrence of a resumptive pronoun seems to correlate with the degree of embeddedness (or, alternatively, the distance between filler and gap): the more deeply embedded the relativized syntactic role (or the longer the distance between filler and gap), the more likely the occurrence of a resumptive pronoun.

This suggests that resumptive pronouns are especially likely to surface in syntactically complex constructions. With respect to L2 Englishes, resumptive pronouns are a common feature of many PCEs such as HKE, IndE (Kortmann, Lunkenheimer & Ehret 2020), NigE, Ghananian English, Cameroon English (Gut 2017: 502), or JamE (Patrick 2004: 427). In the context of L2 acquisition, the use of resumptive pronouns is frequently

attributed to a tendency toward hyperclarity (cf. 2.4.2; Williams 1987: 178). According to Mesthrie & Bhatt (2008: 175), resumptive pronouns “ha[ve] the effect of creating redundancy which is not there in the [target language], as part of the process of clarifying meaning”. As the use of resumptive pronouns thus reflects an effort to reduce ambiguity, it can also be expected to play a role in the present study.

Second-language learning principles can also account for structures with missing or unexpected prepositions. In 2.4.2, it was pointed out that L2 Englishes tend to be characterized by processes of simplification, which, for instance, lead to the omission of redundant prepositions (Gilquin & Granger 2011). In accordance with this assumption, several studies (Bardovi-Harlig 1987: 393; Kao 2001: 203; Sadighi, Parhizgar & Saadat 2004: 22, 25; Suárez-Gómez 2014: 260; Almahammed, Ariff & Sidek 2015: 25) found that learners of English tend to omit stranded and pied-piped prepositions. The use of unexpected prepositions in stranding and pied-piping constructions has not yet been systematically studied. However, the examples provided by Hoffmann (2011: 133) indicate that, just like the novel verb-preposition combinations described in 2.4.2, these innovations might frequently result from a preference for transparent constructions. In (3.100a), for example, *focus at*, which is used instead of the standard variant *focus on*, means ‘give attention to something’. Since the preposition *at* is commonly used as part of direction PPs such as *She looked at him* or *He smiled at the boy*, speakers may consider *focus at*, in which *at* implies movement and directionality, to be more semantically compatible with the meaning ‘direct one’s attention towards something’ than *focus on*.

Another innovation that can be expected to surface in PCEs but which has so far not been mentioned in the literature in connection with L2 Englishes is the use of stranding and pied-piping structures with “extra” (Riley & Parker 1986: 293) prepositions such as (3.104), where *of* is used, while Standard Englishes would omit the preposition.

(3.104) *we’ve got a new product strategy of which you’ll be seeing this this afternoon* <ICE-SIN:S2A-055#59:1:A>

According to Radford (2019: 166), who analyzed this phenomenon in Standard varieties, extra prepositions should be treated as manifestations of “hypercorrection”. Similarly, Liberman (2014) argued that extra prepositions reflect a speaker’s attempt “to ‘sound erudite’ by adding an extra preposition at the start of a relative clause”. Now, considering that New Englishes are frequently characterized by a lack of register distinctions (cf. 2.3), it is doubtful whether speakers of PCEs consciously choose formal variants. However, speakers of New Englishes are usually exposed to formal input, in which pied-piped

prepositions should dominate (cf. 3.1.1.1). Just like the use of structures with doubled or mismatched prepositions, the overuse of pied-piped prepositions in L2 Englishes may thus result from the activation of entrenched P + WH chunks such as *in which* or *of which*. Furthermore, in line with Rohdenburg's (1996: 151) "Complexity principle", which states that "[i]n the case of more or less explicit grammatical options the more explicit one(s) will tend to be favored in cognitively more complex environments", speakers of PCEs may prefer the use of extra prepositions in constructions that are difficult to process. As this is the first study to investigate the use of extra prepositions in filler-gap constructions in PCEs, this type of innovation will be particularly interesting in the corpus analysis.

As the above has shown, the use of innovative preposition placement structures can be attributed to processing factors, second-language learning principles, and the activation of entrenched P + WH combinations and V + P chunks. For the present investigation, it is important to note, though, that not every non-standard token found in a certain ICE component should be considered a conventionalized construction of the respective variety. Drawing on Croft's (2000) distinction between innovation and propagation (cf. 2.4.3.2), Laporte (2021: 332) used the terms "construct" and "construction" to emphasize the importance of distinguishing between the mere creation of innovative form-function mappings and their conventionalization (also see Van Rooy 2011). As she pointed out,

Construction Grammar also offers a theoretical apparatus that highlights the distinction that exists between language production and internalized linguistic conventions, viz. the distinction between (more or less) entrenched 'constructions' and 'constructs'. [...] The majority of features, given their low frequency, are in fact to be situated at the level of 'constructs'. (Laporte 2021: 337)

Following Laporte (2021: 332), it can be expected that a large proportion of the innovative preposition placement tokens identified in the corpus study will be low-frequency phenomena which should be treated as constructs, i.e. as innovations that are not conventionalized at the community level.

3.4 Preposition placement in local languages

In the previous chapters, it has been repeatedly mentioned that innovations in New Englishes often result from cross-linguistic influence. This section thus seeks to give an overview of the preposition placement constructions that exist in the local languages of the PCEs at hand. In the light of the large number of local languages and the scarcity of the literature on preposition placement, this overview does not claim to be exhaustive. Instead, based on a survey of the languages for which it was possible to glean information

on preposition placement, it aims to show which speakers can be expected to receive input for constructions in their L1s⁶ that may trigger interlingual identification (cf. 2.4.3.2). Since WH-interrogatives (cf. 3.4.1) and relatives (cf. 3.4.2) are the clause types for which most information on preposition placement is given in the literature, the main focus will be on these two construction types.

3.4.1 WH-interrogative constructions

A summary of the relevant WH-interrogative constructions that exist in the local languages of the varieties included in the present study is provided in table 3.3. Note that (✓) marks constructions which do not involve actual pre- or postpositions but which share certain properties with stranding and pied-piping constructions in English.

Table 3.3 clearly shows that the majority of the local languages do not license stranding in WH-interrogatives. Instead, they either have pied-piping constructions or in-situ structures, in which the WH-element appears in the position in which the prepositional complement would be found in a declarative construction. For instance, in the English in-situ WH-interrogative structure in (3.105b), the constituent order is the same as in the declarative construction in (3.105a).

- (3.105) (a) *She gave the book to John.*
 (b) *She gave the book to whom?*

While in-situ WH-interrogatives are restricted to “contexts of sustained questioning, such as quizzes and interrogations” (Huddleston 2002c: 873) in English, they are the norm in many local languages of the PCEs at hand. Now, on the one hand, in-situ WH-questions differ from pied-piped interrogatives in that the P + WH element does not appear clause-initially. On the other hand, in-situ structures are more similar to pied-piping than to stranding phenomena since the preposition and the filler are not separated. Consequently, speakers of L1s in which in-situ WH-interrogatives are common might favor pied-piping over stranding in English.

The only variety in which speakers cannot be expected to encounter pied-piped or in-situ WH-interrogatives in their L1 input is JamE. Jamaican Creole, a Caribbean

⁶ Recall that countries in which PCEs are spoken tend to be characterized by multilingual settings in which speakers typically use more than one local language on a daily basis (cf. 2.4.3.1). Consequently, in the present study, the term “L1” does not mean to suggest that speakers acquire only a single local language. Instead, it refers to any local language acquired by a speaker.

English-lexified Creole also known as “Patwa” (Patrick 2004: 407), only has stranded WH-interrogatives:

- (3.106) *A (h)uu im bring dat fa?*
 A who 3SG bring that for
 ‘Who did s/he bring that for?’
 (Durrleman-Tame 2008: 86)

At the same time, pied-piped WH-interrogative constructions exist in major L1s of Singapore, Hong Kong, the Philippines, India and Nigeria, whereas the local languages spoken in Kenya and Tanzania only seem to have in-situ WH-interrogatives.

In Singapore, almost 30% of the population consider Mandarin Chinese the language they most frequently speak at home (Department of Statistics Singapore 2021: 23). Mandarin has preposition-like elements⁷ but it does not license preposition stranding (Zhang 1990: 313). In WH-interrogatives, the WH-word normally remains in-situ (Li & Thompson 1989: 522; Huang, Li & Li 2009: 260). According to Cheung (2008: 138), alternatively, both the interrogative word and the preposition-like element (e.g. *zai nali* in (3.107)) can be pied-piped.

- (3.107) *(Shi) zai nali i Zhangsan i kandao Mali ne?*
 be at where Zhangsan see Mary Q
 ‘Where was it that Zhangsan saw Mary?’
 (example from Cheung 2008: 138)

The other Chinese languages spoken in Singapore, such as Hokkien (Chen 2020: 361) or Cantonese (Matthews & Yip 2011: 373), have in-situ interrogative structures as well. Another important language in Singapore is the Austronesian language Malay, which is the major home language of almost 10% of the population (Department of Statistics Singapore 2021: 23). Just like Mandarin Chinese, Malay prohibits stranding, while it has both pied-piped and in-situ WH-interrogatives (Kader 1981: 179–180; Salleh 1989: 53).

Similar substrate influences can be expected to play a role in HKE. The two official languages of Hong Kong are English and Mandarin Chinese (Setter, Wong & Chan 2010: 5). At the same time, it is Cantonese which most speakers (88.2%) claim to be their usual spoken language (Census and Statistics Department 2022: 45). Pied-piped

⁷ The status of prepositions is a controversial issue in the Chinese languages as the distinction between coverbs and prepositions is not always clear (Li & Thompson 1989: 356–367).

and in-situ structures should thus also be the most common interrogative constructions in the L1s of HKE speakers.

The two official languages of the Philippines are English and Filipino, which is a version of Tagalog enriched by local features of Philippine languages (Kirkpatrick & Lixun 2021: 164). In addition, there are twelve “languages of wider communication”, namely the Austronesian languages Cebuano, Tagalog, Ilokano, Hiligaynon, Bikol, Kapampangan, Pangasinan, Maguindanaon, Tausug, Waray-Waray and Masbatenyo, as well as Mandarin Chinese (Dayag 2016: 103). Since Himmelmann (2005: 142) claims that “no western Austronesian language allows the constituents of the noun phrases or prepositional phrases to be distributed discontinuously across the clause”, these languages do not seem to license preposition stranding either. At the same time, grammatical descriptions of several Philippine languages provide examples of pied-piped WH-interrogative structures involving the preposition *para* (Schachter & Otones 1972: 512–513; Kroeger 1993: 212, 214 on Tagalog; Richards 1971: 268 on Kampampangan; Benton 1971: 87 on Pangasinan; Motus 1971: 239–240 on Hiligaynon). Example (3.108) illustrates this phenomenon for Tagalog:

- (3.108) *Para kanino mo binili ang=laruan?*
 for who(DAT) 2SG.GEN PRF-buy NOM=toy
 ‘For whom did you buy the toy?’
 (adapted from Kroeger 1993: 212; own emphasis)

Moreover, Cebuano has semantic affixes with preposition-like functions that combine with nouns and certain interrogative pronouns (Zorc 1977: 103–104). For instance, in (3.109), the affix *tagá* ‘from’ attaches to the interrogative element *díqín* ‘where’ (Zorc 1977: 107). The resulting structure resembles a pied-piped interrogative construction in English.

- (3.109) *tagà-díqín silá?*
 from-where they
 ‘Where are they from?’
 (adapted from Zorc 1977: 107)

Coming to India, the Indo-Aryan languages, which also include the national language Hindi, account for almost 80% of the country’s L1s (Singh & Nakkeerar 2011: 203). These languages have postpositions that either combine with their objects or follow them as closely as possible (Masica 1991: 213). At the same time, postpositions can

usually not be separated from their complements. For instance, in the SOV-language Hindi, the interrogative element typically remains in-situ and the postposition appears right behind it (Masica 1991: 386–387):

- (3.110) *ye larkiyā kahā se a rahi hāī*
 these girls where from come PROG PRS
 ‘Where are these girls coming from?’
 (adapted from Montaut 2004: 267)

WH-in-situ questions are also the norm in other Indo-Aryan languages such as Marathi (Pandharipande 1997: 16; Dhongde & Wali 2009: 245), Bengali (David 2015: 255), Punjabi (Bhatia 1993: 10, 13) or Oriya (Neukom & Patnaik 2003: 349). At least in some of these languages, the WH-element can also appear in a different position in topicalized interrogative constructions (Montaut 2004: 267; Kachru 2006: 187). In Marathi, for example, the P + WH string (e.g. *kāśā-kāṛta* in (3.111a)) normally precedes the verb, but it can also appear clause-initially (3.111b) or post-verbally (3.111c) (also see Manetta 2011: 138 on Hindi; Wali 2005: 15, 43 on Marathi; Ray 2003 on Oriya). The examples in (3.111b) and (3.111c) can thus be considered instances of postposition pied-piping.

- (3.111) (a) *lili-ni babu-la ratr-i pustāk kāśā-kāṛta dilā?*
 Lili-ERG Babu-DAT night-at book what-for gave
 (b) *kāśā-kāṛta lili-ni babu-la ratr-i pustāk dilā?*
 what-for Lili-ERG Babu-DAT night-at book gave
 (c) *lili-ni babu-la ratr-i pustāk dilā te kāśā-kāṛta?*
 Lili-ERG Babu-DAT night-at book gave that what-for
 ‘For what/why did Lili give Babu a book at night?’
 (adapted from Wali 2005: 43)

In contrast to these SOV-languages, Kashmiri exhibits a verb-second word order in declarative constructions. In interrogatives, the WH-element appears in second position before the predicate (Wali & Koul 1997: 10), which also leads to postposition pied-piping (Wali & Koul 1997: 23).

Table 3.3: Preposition placement in interrogative constructions in local languages

Variety	Language	Stranding	Pied-piping	In-situ	Resumptive
JamE	Jamaican Creole	✓			
SgE	Mandarin		✓	✓	
	Cantonese			✓	
	Hokkien			✓	
	Malay		✓	✓	
HKE	Mandarin		✓	✓	
	Cantonese			✓	
Phile	Filipino/Tagalog		✓		
	Kampampangan		✓		
	Pangasinan		✓		
	Hiligaynon		✓		
	Cebuano		(✓)		
	Mandarin		✓	✓	
IndE	Hindi		✓	✓	
	Punjabi			✓	
	Bengali			✓	
	Marathi		✓	✓	
	Oriya		✓	✓	
	Kashmiri		✓		
	Tamil			✓	
	Telugu			✓	
	Malayalam			✓	
	Kannada			✓	
NigE	Hausa	(✓)	✓		
	Igbo		✓	✓	
	Yoruba	✓	✓		
	Kanuri		✓	✓	
	Ibibio		✓	✓	
	Tiv	✓	✓		✓
	Igala				✓
	Nigerian Pidgin	(✓)	✓	✓	
KenE	Swahili			✓	
	Kamba			✓	
	Dholuo			✓	
	Pokot			✓	
TanE	Swahili			✓	
	Haya			✓	
	Dholuo			✓	

The second important language group in India is the Dravidian language family, which makes up about 20% of all local languages (Singh & Nakkeerar 2011: 203). The Dravidian languages, which include e.g. Tamil, Telugu, Malayalam and Kannada, also have postpositions and in-situ WH-interrogatives (Asher 1982: 10 on Tamil; Arden 1975: 133 on Telugu; Asher & Kumari 1997: 7, 14 on Malayalam) although the WH-word can also appear in other positions for stylistic reasons (see e.g. Asher & Kumari 1997: 21 on Malayalam; Sridhar 1990: 8 on Kannada).

Nigeria is estimated to have between 395 and 526 local languages, with the three “majors” being Hausa, Yoruba, and Igbo (Jowitt 2019: 5). Both the Afro-Asiatic language Hausa and the Niger-Congo language Igbo permit pied-piping in interrogatives (Newman 2000: 471 on Hausa; Amaechi 2020: 49 on Igbo), while only Igbo has in-situ WH-questions (Amaechi & Georgi 2019: 17–19). Furthermore, even though both languages generally prohibit stranding (Newman 2000: 471 on Hausa; Green & Igwe 1963: 45; Emenanjo 2015: 22 on Igbo), Hausa has structures involving the indirect object marker *wà* which resemble stranded WH-interrogative constructions in English. As shown in (3.112), *wà* functions like a preposition and can be stranded freely (Newman 2000: 276–277).

- (3.112) *wane ne za ka nuna wà hoton?*
 Who is it will you show to picture
 ‘Who are you going to show the picture to?’
 (adapted from Newman 2000: 277)

In Yoruba, another Niger-Congo language, the situation is slightly more complicated. Depending on the type of preposition involved, Yoruba permits either stranding or pied-piping. As (3.113) illustrates, the preposition *sí* ‘to’ can be stranded, while the preposition *ni* ‘at’ only licenses pied-piping (Adesola 2006: 10; Amaechi & Arokoyo 2022: 90).

- (3.113) (a) *Kí ni Olú da omi sí*
 what be Olu pour water to
 ‘What did Olu pour water into?’
 (b) * *Ibo ni Olú wà ní*
 where be Olu exist at
 ‘Where is Olu?’
 (Adesola 2006: 10; own emphasis)

In addition, there are the so-called “major minors” with more than one million L1 speakers each, namely Fulfulde, Kanuri, Ibibio, Tiv, Ijo, Ebira, Igala, Nupe, and Berom (Jowitt 2019: 5). With respect to preposition placement in interrogatives, the literature provides information on Kanuri, Ibibio, Tiv and Igala.

Kanuri has postpositional clitics which cannot stand by themselves but always have to attach to an object (Fannami 2011: 112–113). In interrogatives, the WH-element can be placed clause-initially or in-situ, but the postposition is always attached to the interrogative word (Fannami 2011: 57–58). Ibibio also licenses both pied-piped and in-situ WH-questions (Essien 1990: 159–160). In Tiv, the choice between stranded and pied-piped WH-interrogatives is largely determined by the preposition type. Prepositions which have to be followed by an obligatory complement, such as *shá* ‘on’ or *shin* ‘in, from’, cannot be stranded (3.114). In contrast, prepositions which do not require an obligatory complement, such as *ké* ‘in’ (3.115), appear in both stranding and pied-piping constructions. Alternatively, they can also be followed by a resumptive pronoun (e.g. *mín* in (3.115)), which avoids stranding (Angitso 2014: 52–54).

- (3.114) * *Nyí Sésùgh á úndé shá?*
 what Sésùgh AGR.SG.PST climb-PST on
 ‘What did Sésùgh climb on?’

- (3.115) *Nyí Sésùgh á nyôr ké (mín)?*
 what Sésùgh AGR.SG.PST enter-PST in it
 ‘What did Sésùgh enter into?’

(adapted from Angitso 2014: 53)

Structures in which a pronominal clitic attaches to a preposition that would otherwise be stranded are also found in Igala (Ejeba 2016: 158).

Another language that should be mentioned in the context of Nigeria is Nigerian Pidgin, i.e. a dialect of Afro-Caribbean English-lexifier Creole (Faraclas 2008: 340). Although particularly educated Nigerians “prefer not to speak it because of its demotic, anti-bourgeois connotations” (Jowitt 2019: 10), Nigeria Pidgin is frequently considered a lingua franca in the South of Nigeria (Jowitt 2019: 11). With respect to preposition placement, Nigerian Pidgin has in-situ WH-interrogatives (3.116a), but prepositions can also be pied-piped to the front of the clause (Faraclas 1996: 12, 16). In contrast, preposition stranding is normally prohibited. The only element that can be stranded in interrogatives is the element *fròm* (3.116b), which “often behaves more like a serialized verb than a preposition” (Faraclas 1996: 16).

- (3.116) (a) *Yù kom fròm we?*
 2SG come from where
 (b) *We yù kom fròm?*
 where 2SG come from
 ‘Where do you come from?’

(adapted from Faraclas 1996: 16)

The fact that Yoruba and Tiv (and, at least to a certain extent, Hausa and Nigerian Pidgin) license stranding in WH-interrogatives indicates that speakers of NigE should be more likely to receive L1 input for interrogative structures with stranded prepositions than speakers of the other varieties discussed so far (the only exception being JamE). Consequently, speakers of NigE may show a stronger preference for stranded WH-interrogative constructions than speakers of other varieties such as IndE or PhilE who should not be exposed to stranded WH-interrogative structures in their L1s at all.

While all of the varieties mentioned so far have major local languages which permit pied-piping in WH-interrogatives, the L1s spoken in Kenya and Tanzania only seem to license in-situ WH-questions. 65% of the languages spoken in Kenya belong to the Bantu languages (Musau 2004: 60; Buregeya 2019: 15). Swahili, which is Kenya’s lingua franca and, in addition to English, an official language of the country (Kamwangamalu & Tovares 2016: 424; Buregeya 2019: 9), is also a Bantu language. Swahili does not license preposition stranding, but in interrogatives the WH + P combination appears in-situ (Krifka 1995: 1415; Thompson & Schleicher 2001: 75–77). At the same time, the Bantu languages only have a small number of prepositions, while they frequently make use of clitics with prepositional meanings that combine with other elements (Riedel 2010: 218). For instance, Kamba has complex interrogative forms such as *naũũ* ‘with whom’ that consist of a WH-element (e.g. *ũũ* ‘who’) and a prefix with a prepositional meaning (e.g. *na-* ‘with’; Whiteley & Muli 1962: 98). Hoffmann (2011: 106–107) suggested that these structures “might potentially be seen as cases of pied-piping, which would mean that L1 transfer should favour this preposition placement variant”.

The Nilotic languages, which account for about 30% of all Kenyan L1s, make up another important language group in Kenya (Musau 2004: 60; Buregeya 2019: 15). According to the literature, in-situ P + WH interrogatives exist in Pokot, a dialect group of the South Nilotic Kalenjin languages (Baroja, Sikamoy & Partany 1989: 95–96), and in the Western Nilotic language Dholuo (Omondi 1982: 142; Okoth-Okombo 1997:

1997). This indicates that in-situ WH-interrogatives might be the norm in the Nilotic languages.

In addition, about 3% of all Kenyan languages belong to the Cushitic language family (Musau 2004: 60; Buregeya 2019: 15). However, it was not possible to find information on preposition placement in these languages. Hoffmann (2011: 107) speculated that it might be relevant that at least some Cushitic languages, such as Boni (Heine 1982: 54, 67) or Oromo (Griefenow-Mewis 2001: 50–53), make use of clitics to express prepositional meanings. Since clitics can usually not stand by themselves, these languages might pattern with the Bantu languages in terms of preposition placement, i.e. they might avoid preposition stranding (Hoffmann 2011: 107).

The linguistic landscape in Tanzania is very similar to that of Kenya. Just like in Kenya, the dominant local language is Swahili, which is spoken as a first language by about 10% and as a second language by about 90% of the population. Other major Bantu languages are Kisukuma (12.5 %) and Kinyambwezi (4.2%; Webb & Kembo-Sure 2008: 51). As a result, it is not surprising that it is assumed that "[v]irtually all of Tanzania's inhabitants speak Bantu languages" (Tanzania Government 2015). The literature on preposition placement in these languages is scarce. The only Tanzanian Bantu language for which explicit information on preposition placement is provided is Haya, which has in-situ interrogative structures such as (3.117) (Bennett 1977: 174, 178).

- (3.117) *a-ka-gi-shál-a ná kī*
 he-3SG-it-cut with what
 ‘What did he cut it with?’
 (Bennett 1977: 178; own emphasis)

However, as the Bantu languages of Tanzania can be expected to behave similarly to the Bantu languages of Kenya, they should also be characterized by a preference for in-situ WH-interrogatives and a general tendency to use clitics.

The other local languages spoken in Tanzania belong to the Nilotic, Cushitic and Khoisan languages (Webb & Kembo-Sure 2008: 51). While the WH-in-situ structures of the Nilotic language Dholuo have already been mentioned in the context of Kenya, the other Tanzanian languages have only very few speakers (e.g. the Khoisan languages are only spoken by 1000 people each on average; see Webb & Kembo-Sure 2008: 31) and are hardly dealt with in the literature. However, considering the parallels between the substrate systems of TanE and KenE, these languages should also tend to avoid

preposition stranding, while in-situ structures can be assumed to be the most common WH-interrogative type.

With respect to the present study, this sketch of L1 interrogative constructions allows the following conclusions:

Due to the stranding requirement of Jamaican Creole, JamE should be the variety in which speakers receive the most L1 input for stranded WH-interrogatives. At the same time, Jamaican Creole does not have pied-piped or in-situ WH-questions. Consequently, speakers of JamE may favor stranded WH-interrogative constructions more than speakers of other varieties.

At least some speakers of NigE (e.g. speakers of Yoruba or Tiv) should also be exposed to stranded WH-interrogatives in local languages. Speakers of NigE may thus also have a stronger preference for stranded WH-interrogative constructions than speakers of PCEs that only have L1s which do not license stranding at all.

The local languages of SgE, HKE, PhilE, IndE, KenE and TanE seem to prohibit stranding in interrogatives. Since speakers of these varieties should encounter a lot of L1 input for pied-piped and in-situ WH-interrogative structures, cross-linguistic influence may lead to a pied-piping preference.

3.4.2 *Relative constructions*

While pied-piped and P + WH in-situ structures were identified as the dominating interrogative types in most local languages, the situation is more complex in relative clauses. Table 3.4 gives an overview of all relevant relative constructions that exist in the local languages of the PCEs at hand. Most strikingly, there are only two varieties that have major L1s in which speakers can be expected to receive input for pied-piped relative clause constructions, namely IndE and NigE.

With respect to India, it is important to remember that the Indo-Aryan languages have postpositions that cannot be separated from their complements. In relative clauses, a postposition thus normally appears right behind the relative pronoun (see Keenan & Comrie 1979: 338 on Hindi; Bhatia 1993: 59 on Punjabi; Pandharipande 1997: 88 on Marathi; Wali & Koul 1997: 54–61 on Kashmiri; David 2015: 97 on Bengali). Although this phenomenon is not referred to as pied-piping in the literature, structures such as the Hindi example in (3.118), in which the postposition *se* follows the relativizer *jis*, can be considered cases of postposition pied-piping.

- (3.118) *Ram ne us caakuu ko dekhaa jis se us*
 Ram SBJ that knife OBJ saw REL INS that
aadmii ne murgii ko maaraa thaa
 man SBJ chicken OBJ killed
 ‘Ram saw the knife with which the man killed the chicken.’
 (adapted from Keenan & Comrie 1979: 338)

The Dravidian languages prefer participial relative clause without an overt relative pronoun, in which relativization is signaled by an affix that attaches to the verb (Asher & Kumari 1997: 53; Hendery 2012: 170). The Malayalam example in (3.119) shows that postpositions are normally lost in these structures (see Asher & Kumari 1997: 58–68; Sreekumar 2020: 157 on Malayalam; Krishnamurti 1985: 247 on Telugu; Sridhar 1990: 56–58 on Kannada). While the noun *vati* is followed by the postposition *konta* in the declarative construction in (3.119a), the postposition does not form part of the relative clause in (3.119b).

- (3.119) (a) *pooliiskaaran vati konta kuttiye aticcu*
 policeman stick-INSTR child-ACC beat-PST
 ‘The policeman beat the child with a stick.’
 (b) *pooliiskaaran kuttiye aticca vati*
 Policeman child-ACC beat-PST-REL stick
 ‘the stick with which the policeman beat the child’
 (adapted from Asher & Kumari 1997: 60)

At the same time, Malayalam (Asher & Kumari 1997: 64, 122), Tamil (Asher 1982: 30, 60; Lehmann 1989: 292) and Kannada (Sridhar 1990: 50) license stranding in participial relative clauses with a small number of locative postpositions. As Lehmann (1989: 292) points out for Tamil, “NPs occurring in a postpositional phrase with a free postposition can be relativized, leaving a postposition without a noun phrase argument behind”. While usage guides do not refer to these structures as cases of postposition stranding, examples such as (3.120), a relative construction from Malayalam, indicate that speakers of certain Dravidian languages should also encounter input for relative constructions in which postpositions are separated from their complements (e.g. *miite* does not appear behind *mesa* in (3.120b)).

Table 3.4: Preposition placement in relative constructions in local languages⁸

Variety	Language	Stranding	Pied-piping	Resumptive	No prepositional relatives
JamE	Jamaican Creole	✓			
SgE	Mandarin			✓	
	Cantonese			✓	
	Hokkien			✓	
	Malay		✓		✓
HKE	Mandarin			✓	
	Cantonese			✓	
Phile	Filipino/Tagalog				✓
	Kampampangan				✓
	Mandarin			✓	
IndE	Hindi		✓		
	Punjabi		✓		
	Bengali		✓		
	Marathi		✓		
	Kashmiri		✓		
	Telugu	✓			
	Malayalam	✓			
	Kannada	✓			
	Tamil	✓			
NigE	Hausa		✓	✓	
	Igbo			✓	
	Yoruba	✓	✓		
	Nupe	✓			
	Kanuri			✓	
	Nigerian Pidgin			✓	
KenE	Swahili			✓	
	Luyia			✓	
	Dholuo	✓			
	Turkana	✓			

⁸ According to Himmelmann (2005: 161), the ban on prepositional relatives in the Philippine languages should apply to all Western Austronesian languages, i.e. not just to Tagalog and Kampampangan but also to Cebuano, Ilokano, Hiligaynon, Bikol, Pangasinan, Maguindanaon, Tausug, Waray-Waray and Masbatenyo. However, since other grammatical descriptions only explicitly mention this restriction for Tagalog and Kampampangan, table 3.4 only includes these two languages.

Table 3.4 (cont.)

Variety	Language	Stranding	Pied-piping	Resumptive	No prepositional relatives
TanE	Swahili			✓	
	Haya			✓	
	Dholuo	✓			

(3.120) (a) *naan meesayute miite irunnu*

I table-GEN upon sit-PST

‘I sat on the table’

(b) *naan miite irunna mesa*

I upon sit-PST-REL table

‘the table that I sat on’

(adapted from Asher & Kumari 1997: 64)

However, taking into account that the Indo-Aryan languages make up a much larger proportion of the languages of India than the Dravidian languages (78% vs. 20%; Singh & Nakkeerar 2011: 203), for most speakers of IndE, cross-linguistic influence should lead to a preference for pied-piping.

The only other variety with major local languages that license pied-piping in relative clauses is NigE. In Nigeria, pied-piped relatives can be found in Hausa and Yoruba. In Hausa, prepositions either have to be pied-piped or a resumptive pronoun has to be added to avoid preposition stranding (Newman 2000: 471–472). In Yoruba, it is again the preposition itself that determines preposition placement. Stranding is possible with *fì*, *tí*, *bá*, *fún* and *sí*, while the preposition *ní* must either be pied-piped or omitted (Awobuluyi 1982: 95).

With respect to Igbo, Amaechi (2020: 29) shows that a resumptive pronoun is required when an object of the preposition *màkà* ‘about’ is relativized. For instance, in (3.121), *màkà* cannot be stranded but must be followed by the pronoun *yà*.

(3.121) *Àdà mà ónyé ányí nà-èkwú màkà yà.*

Ada know person 1PL IPFV-NMZL.talk about 3SG

‘Ada knows the person that we are talking about.’

(adapted from Amaechi 2020: 29)

Although there is no further information on preposition placement in Igbo relative clauses, this might be taken as an indication that prepositional relative clauses with resumptive pronouns are common in this language.

Turning to the other languages spoken in Nigeria, the literature on preposition placement in relative clauses is scarce. According to Kandybowicz (2008: 45), stranded relative clause constructions exist in Nupe. In contrast, Kanuri requires resumptive pronouns in prepositional relatives (Hutchison 1981: 220–222). As Haspelmath & Michaelis (2013: 376) point out, resumptive pronouns are also used in instrument relative clauses in Nigerian Pidgin. Compared to the local languages of India, in which pied-piped relatives should be dominant due to the influence of the Indo-Aryan languages, the L1s of Nigeria thus have a greater variety of stranding, pied-piping and resumptive structures. As a result, it is difficult to identify a single relative type that might have an effect on preposition placement in NigE.

Relative constructions with resumptive pronouns can also be found in the Bantu languages of Kenya and Tanzania. In Swahili, for example, a pronominal clitic (e.g. *-ye* in (3.122)) has to be suffixed to the preposition if a prepositional object is relativized (Vitale 1981: 89; Barrett-Keach 1985: 48; Thompson & Schleicher 2001: 102; also see Appleby 1961: 18 on Luyia; Riedel 2010: 218 on the Tanzanian language Haya).

- (3.122) *mut amba-ye Watoto wa-li-pony-w-a na-ye ni*
 man PRO-REL children they-PST-cure-PASS by-REL is
mganga
 medicine man
 ‘the man who the children were cured by is the medicine man’
 (Vitale 1981: 96; own emphasis)

In contrast, stranding is possible in at least two Nilotic languages. Example (3.123) shows that relative clauses with stranded prepositions exist in Dholuo, a language spoken in both Kenya and Tanzania (Omondi 1982: 185).

- (3.123) *puodhó má néné wacchúoyo .e odumâ*
 garden which long ago we-planted in maize
 ‘the garden which we planted maize in long ago’
 (adapted from Omondi 1982: 247; cited in Hoffmann 2011: 107)

Moreover, Turkana, an Eastern Nilotic language spoken in Kenya (Webb & Kembo-Sure 2008: 31), seems to have stranded locative relative constructions (Dimmendaal 1983: 313).

With respect to the Cushitic languages, no information on preposition placement in relative clauses exists. However, what has been pointed out in connection with WH-interrogatives (cf. 3.4.1) should also apply to relatives. Since the Cushitic languages typically use clitics instead of free prepositions, these languages might avoid stranding (Hoffmann 2011: 107).

The major L1s spoken in Singapore do not have stranded relative clause constructions either. As illustrated in (3.124), in Mandarin Chinese elements with preposition-like functions are normally lost under relativization (Li & Thompson 1989: 583).

- (3.124) *páshǒu tōu dōngxī de fāngfǎ*
 pickpocket steal thing REL method
 ‘the method by which pickpockets steal things’
 (Li & Thompson 1989: 583)

Li & Thompson (1989: 584) give some examples of relative clauses in which prepositions are retained, but these are "awkward to many speakers" and "rarely found in speech or writing" (Li & Thompson 1989: 585). In addition, the use of resumptive pronouns is obligatory in these structures. Other Chinese languages spoken in Singapore such as Cantonese (Matthews & Yip 2011: 331) or Hokkien (Chen 2020: 442) also require resumptive pronouns in prepositional relative clauses.

In Malay, another important L1 in Singapore, the relativization of prepositional objects is usually not possible at all. Relative clauses are normally introduced by the invariable particle *yang*, but this strategy cannot be used to relativize prepositional objects (Keenan & Comrie 1977: 71). According to Keenan & Comrie (1977: 71), the only exception is a relative construction that is introduced by a pied-piped preposition followed by the interrogative pronoun *siapa* (e.g. *kapada siapa* in 3.125). However, as these structures are “felt to be bookish and clumsy” (Keenan & Comrie 1977: 71), they should occur rarely in actual language use.

- (3.125) *perempuan kapada siapa Ali beri ubi kentang itu*
 woman to who Ali give potato the
 ‘the woman to whom Ali gave the potato’
 (Keenan & Comrie 1977: 71; own emphasis)

Speakers of SgE should thus mostly receive L1 input for relative constructions in which prepositions are omitted or in which the use of a resumptive pronouns avoids stranding. As the substrate systems of SgE and HKE are very similar (cf. 3.4.1), the same should also apply to speakers of HKE.

The Austronesian languages spoken in the Philippines do not license prepositional relatives either. According to Himmelmann (2005: 161), “Western Austronesian languages are famous for their restrictions on the formation of relative clauses [...]. The basic and central observation is that among the core arguments of a predicate, only subjects may be relativized”. Consequently, none of the Philippine languages mentioned in section 3.4.1 should have relative clauses in which the complement of a PP is relativized (also see Keenan & Comrie 1979: 347; Kroeger 1993: 24; Law 2016: 747 on Tagalog; Richards 1971: 189 on Kampampangan). This effect is also illustrated in the relative example in (3.126), which would be considered ungrammatical in Tagalog.

- (3.126) **si Pedro ang bata-ng i-b<in>igay ng*
 ABS Pedro ABS child-LK APP-<PRF>give ERG
babae sa ang manga.
 woman P ABS mango
 ‘Pedro is the child who a woman gave the mango to.’
 (adapted from Law 2016: 747)

The only language of wider communication in the Philippines that is not an Austronesian language is Mandarin Chinese (cf. 3.4.1), in which prepositions are normally omitted in relatives. As a result, speakers of PhilE should hardly be exposed to stranded or pied-piped relative clause constructions in local languages.

Finally, with respect to JamE, the same picture emerges as in WH-interrogatives. Jamaican Creole licenses stranding (3.127) but prohibits pied-piping in relative clauses. According to Patrick (2004: 426), this can explained by the fact that “prepositions and other post-verbal particles are tightly bound to the verb” in Jamaican Creole.

- (3.127) *di dort we shi waak pan*
 the groundREL she walk on
 ‘the ground which she walks on’
 (adapted from Patrick 2004: 426)

In sum, based on the above, the following cross-linguistic influences can be expected in the present study:

Due to the importance of the Indo-Aryan languages, IndE can be considered the variety in which speakers receive the most L1 input for pied-piped relative structures. Consequently, IndE should have the strongest preference for pied-piped relative clause constructions.

The only other variety in which cross-linguistic influence might lead to a pied-piping preference is NigE. However, considering that NigE also has L1s with stranded or resumptive relative structures, the pied-piping preference of NigE may be weaker than that of IndE.⁹

In contrast, speakers of JamE may have a stronger preference for stranded relative constructions than speakers of all other varieties, due to the stranding requirement of Jamaican Creole.

As a result of cross-linguistic influence, several varieties (SgE, HKE, NigE, KenE, TanE) may favor the use of resumptive pronouns. The only varieties that do not have local languages with resumptive prepositional structures are JamE, IndE and PhilE (disregarding the influence of Mandarin). However, as resumptive pronouns are a common feature of L2 Englishes (cf. 2.4.2), it will be difficult to disentangle cross-linguistic influence from L2 acquisition effects.

The same also applies to structures with missing prepositions. The omission of prepositions may be particularly frequent in PhilE, whose local languages do not have prepositional relatives, and in SgE and HKE since the local languages of these varieties tend to omit prepositions in relative constructions. At the same time, the omission of prepositions is also a general simplification strategy that is likely to affect all PCEs (cf. 2.4.2).

This chapter introduced various factors that are assumed to influence the competition between preposition stranding and pied-piping in English. Section 3.1 demonstrated that

⁹ At first sight, table 3.4 may not necessarily seem to support this claim. However, the Indo-Aryan languages, which account for almost 80% of all Indian languages (cf. 3.4.1), are clearly the most influential language family in India. In contrast, in Nigeria, Hausa, Igbo and Yoruba all play a key role, “each being the principal L1 in several of the country’s states” (Jowitt 2019: 5). While speakers of IndE should thus mostly encounter pied-piped relatives in the Indo-Aryan languages, speakers of NigE are more likely to receive L1 input for stranded and resumptive relative structures.

the usage-based construction network of BrE can be said to contain various stranded and pied-piped macro- and meso-constructions that are interconnected by vertical and horizontal relations. 3.2 showed that the choice between preposition placement constructions is strongly driven by processing factors: While pied-piping is particularly associated with the cognitively most demanding environments, such as relative clauses, NP-contained PPs or passive constructions, stranding is preferred when the verb and the preposition are closely associated with each other. 3.3 focused on innovative uses of preposition placement constructions in L2 Englishes, which are frequently attributed to input effects, processing constraints, cross-linguistic influence and second-language learning principles. Finally, section 3.4 pointed out ways in which cross-linguistic influences may affect the competition between stranding and pied-piping in the eight PCEs at hand. All these factors will have to be taken into account in the present analysis. The next chapter will provide the methodological basis for the corpus study.

4. Data and methods

This chapter will focus on the data and methodology used in the present study. 4.1 will describe the International Corpus of English, which serves as the database for the analysis. 4.2 will outline how the relevant preposition placement tokens were extracted, while 4.3 will focus on the data annotation process. At the end of the chapter (4.4), three statistical techniques will be introduced which were used to analyze the corpus data.¹⁰

4.1 The International Corpus of English

The data for the present study come from the International Corpus of English (ICE). The ICE project was initiated in the late 1980s in order to provide researchers with a database for comparative studies in the field of World Englishes (Greenbaum & Nelson 1996: 3). To date, researchers have completed or are in the process of compiling ICE components for 26 varieties.¹¹ This study includes the 12 ICE varieties for which spoken and written data is available, namely ICE-Great Britain (ICE-GB), ICE-Canada (ICE-CAN), ICE-Ireland (ICE-IRE), ICE-New Zealand (ICE-NZ), ICE-Jamaica (ICE-JA), ICE-Singapore (ICE-SIN), ICE-Hong Kong (ICE-HK), ICE-India (ICE-IND), ICE-Nigeria (ICE-NIG), ICE-Philippines (ICE-PHI), and ICE-East Africa (ICE-EA), which consists of ICE-Kenya (ICE-EA-K) and ICE-Tanzania (ICE-EA-T).

Each ICE component comprises 500 texts consisting of 2,000 words each, which amounts to a total size of approximately one-million words per corpus (Greenbaum & Nelson 1996: 5). Now, in the age of electronic corpora, the ICE project can, of course, be considered fairly small (Nelson 2017: 368–369). For instance, the corpus of Global Web-Based English (GloWbE), which consists of 1.9 billion words scraped from webpages¹², offers much higher token frequencies, which is important particularly for studies investigating infrequent constructions. However, as Nelson (2017: 369) pointed out, “[w]e should not [...] rush to the conclusion that bigger is necessarily better in all areas of linguistic enquiry”. First of all, preposition placement is a fairly frequent syntactic phenomenon for which the size of the ICE corpora should be sufficient (as shown by e.g. Hoffmann’s (2011) analysis of preposition placement based on ICE-GB and ICE-EA-K).

¹⁰ Data and R code of the entire analysis are available at https://osf.io/y6zth/?view_only=1fb185fd8cbc4c138e0b2f247f0d0686.

¹¹ According to the ICE website: <https://www.ice-corpora.uzh.ch/en.html> (13 September, 2023)

¹² According to the GloWbE website: <https://www.english-corpora.org/glowbe/> (13 September, 2023)

Moreover, the ICE corpora have the major advantage that they all follow a common design which includes a variety of text types from spoken and written registers (cf. table 4.1). Preposition placement being strongly influenced by processing factors (cf. 3.2), it was paramount to use a database that includes spoken language, in which processing effects are most likely to surface due to the high processing load associated with real-time language production (Biber 1988: 163). Furthermore, preposition placement is also affected by stylistic variables (cf. 3.1.1.1). Since the ICE corpora cover both formal (e.g. academic writing or legal presentations) and informal registers (e.g. private dialogues), they are ideal for an investigation of formality effects. Despite their rather small sample size, the ICE corpora were thus preferred over a larger database such as GloWbE, which only includes written language.

Nevertheless, the ICE corpora have a few limitations which should be noted briefly. The first issue concerns the comparability of the corpora. Although the various ICE components generally follow the same corpus design, there is one major exception, namely ICE-EA. To account for the different sociolinguistic situations in Kenya and Tanzania, ICE-EA consists of two individual subcomponents for KenE and TanE (Hudson-Ettle & Schmied 1999: 4). In addition, in both Kenya and Tanzania, it turned out difficult to collect spoken data. As a result, the private dialogue section of ICE-EA is much smaller than that of the other ICE components and mostly restricted to Kenyan texts. While the ten other ICE components at hand all have a private dialogue category consisting of 100 texts (90 direct and 10 telephone conversations), the corresponding section in ICE-EA comprises only 30 texts from KenE and three from TanE, with phone conversations not being included at all. At the same time, an additional category called “broadcast lessons” was added to the Kenyan data to make up for the lack of private conversation data (Hudson-Ettle & Schmied 1999: 6–7). Moreover, in both Kenya and Tanzania, it was impossible to obtain data for the categories “parliamentary debates” and “legal cross-examinations”. Instead, an additional category called “written as spoken” was introduced, which consists of Hansards and written recordings of cross-examinations from Kenya (Hudson-Ettle & Schmied 1999: 7–8). In the present study, it should thus be kept in mind that the structure of ICE-EA differs from that of the other ICE components, and that particularly ICE-EA-T hardly includes informal, spoken data.

Although the other ICE corpora follow the design shown in table 4.1, they are not perfectly comparable either. For instance, in ICE-IRE, the legal texts do not come from courtrooms but from legislative committee hearings because courtroom recordings are

Table 4.1: ICE text categories (based on Nelson 2002: 4–5)¹³

Category		File name	
SPOKEN			
DIALOGUE	PRIVATE	Direct conversations	S1A-001 to S1A-090
		Telephone calls	S1A-091 to S1A-0100
	PUBLIC	Class lessons	S1B-001 to S1B-020
		Broadcast discussions	S1B-021 to S1B-040
		Broadcast interviews	S1B-041 to S1B-050
		Parliamentary debates	S1B-051 to S1B-060
		Legal cross-examinations	S1B-061 to S1B-070
		Business transactions	S1B-071 to S1B-080
MONOLOGUE	UNSCRIPTED	Spontaneous commentaries	S2A-001 to S2A-020
		Unscripted speeches	S2A-021 to S2A-050
		Demonstrations	S2A-051 to S2A-060
		Legal presentations	S2A-061 to S2A-070
	SCRIPTED	Broadcast news	S2B-001 to S2B-020
		Broadcast talks	S2B-021 to S2B-040
		Non-broadcast talks	S2B-041 to S2B-050
WRITTEN			
NON-PRINTED	NON-PROFESSIONAL	Student essays	W1A-001 to W1A-010
		Examination scripts	W1A-011 to W1A-020
PRINTED	CORRESPONDENCE	Social letters	W1B-001 to W1B-015
		Business letters	W1B-016 to W1B-030
	ACADEMIC WRITING	Humanities	W2A-001 to W2A-010
		Social sciences	W2A-011 to W2A-020
		Natural sciences	W2A-021 to W2A-030
		Technology	W2A-031 to W2A-040
	NON-ACADEMIC WRITING	Humanities	W2B-001 to W2B-010
		Social sciences	W2B-011 to W2B-020
		Natural sciences	W2B-021 to W2B-030
		Technology	W2B-031 to W2B-040
	REPORTAGE	Press news reports	W2C-001 to W2C-020
	INSTRUCTIONAL WRITING	Administrative writing	W2D-001 to W2D-010
		Skills & hobbies	W2D-011 to W2D-020
	PERSUASIVE WRITING	Press editorials	W2E-001 to W2E-010
	CREATIVE WRITING	Novels & stories	W2F-001 to W2F-020

¹³ ICE-EA and ICE-NIG use different file names.

prohibited in Ireland (Kallen & Kirk 2008: 9). Another example that shows that the comparability of the corpora should not be taken for granted is the private dialogue section of ICE-HK. Since Cantonese speakers of Hong Kong typically only use English if a non-Cantonese speaker is present, the private conversations of ICE-HK always required the participation of an additional non-Cantonese speaking person (Bolt & Bolton 1996). Consequently, the private dialogues of ICE-HK may not be as natural and as informal as those of the other corpora. For the interpretation of the results, it will thus be important to remember that differences between varieties may sometimes also be due to characteristics of the data itself.

Moreover, the individual ICE components differ from each other with respect to the time period in which the data were collected. While the first corpora, namely ICE-GB, ICE-SIN and ICE-IND were compiled between 1990 and 1994 (Nelson 1996: 28), the data collection for the other components lasted much longer, i.e. until 1996 for ICE-EA (Hudson-Ettle & Schmied 1999: 5) and ICE-HK (Bolt & Bolton 1996: 198), until 1998 for ICE-NZ (Vine 1999: 8), until 2003 for ICE-IRE (Kallen & Kirk 2008: 5), until 2004 for ICE-PHI (Bautista 2004), until 2005 for ICE-JA (Rosenfelder et al. 2009: 1), until 2009 for ICE-CAN (Newman & Columbus 2010: 4), and until 2014 for ICE-NIG (Gut 2014: 2). However, as the present study focuses on preposition placement, i.e. a syntactic phenomenon that should be fairly stable, these minor time difference should not be problematic.

Finally, it should be noted that the ICE corpora represent educated English. The corpora typically include language produced by “adults (over 18) who have received formal education through the medium of English to the completion of secondary level schooling” (Greenbaum & Nelson 1996: 5). Hence, the results of the present study will only allow conclusions about the educated forms of the varieties.

4.2 Data preparation and extraction

4.2.1 Corpus preparation

The study used the raw text files of the ICE corpora. However, before the data could be extracted, a few additional steps were necessary to prepare and clean the corpora.

To ensure that no irrelevant material was included in the analysis, all extra-corpus material, i.e. language material not produced by corpus speakers, had to be removed. Although the beginning and the end of extra-corpus material is marked by tags (<X> and </X>), these tags are sometimes used incorrectly. In order to identify such markup errors,

for each corpus, the number of opening and closing tags was determined running regular expressions in the text editor *Sublime Text*. If the number of opening and closing tags in a corpus did not match, errors were manually inspected, missing tags were added and mistakes were corrected (e.g. the use of `<X/>` instead of `</X>`). Moreover, the opening tags in ICE-EA (`<X_>`) were replaced with the general ICE version (`<X>`), which made it possible to use the same regular expressions in all of the corpora. After these modifications, the extra-corpus material was excluded with the help of an *R* script¹⁴ (R Core Team 2022).

Two more preparatory steps had to be taken in order to facilitate the data extraction (cf. 4.2.2). The ICE corpora normally include text unit markers such as `<ICE-GB:S1A-001 #7:1:A>` which provide information about the variety (e.g. ICE-GB), the file name and the category (e.g. S1A-001), the text unit (e.g. 7), the subtext (e.g. 1), and the speaker (e.g. A) to which an utterance belongs. As the study relied on automatic searching (cf. 4.2.2), these markers were of crucial importance for the correct identification of the extracted tokens. Regular expressions were therefore also used to identify mistakes in text unit markers (e.g. in the file S1A-060 of ICE-CAN, the string `ICE-CAN:S1A-060` is missing). Moreover, with the help of the *Find* and *Replace* functions of *Sublime Text*, *ICE-EA*- and *ICE-IRE*-markers were added to the files of ICE-EA and ICE-IRE¹⁵, which do not always follow the general ICE markup scheme (e.g. ICE-EA simply uses `<#/>` to mark a new text unit).

Finally, ICE-EA differs from the other corpora in that all texts of the same category are combined into a single file. To make sure that all corpora could be processed and searched in the same way, an *R* script was thus used to split the large ICE-EA files into individual text files.

4.2.2 Data extraction

The quickest way to extract preposition placement tokens from the corpora would, of course, be a fully automatic approach. For example, in part-of-speech (POS) tagged versions of the corpora, pied-piped prepositions could be identified by searching for P + WH sequences, whereas stranded prepositions could be extracted by looking for prepositions followed by punctuation marks. However, such an automatic approach is

¹⁴ I would like to thank Thomas Brunner for his support.

¹⁵ As the file names of ICE-NIG differ from all other ICE components, *ICE-NIG*-tags were created in a later step with the help of the *R* script that was written to extract all WH-tokens (cf. 4.2.2).

highly unreliable as it would miss all stranded prepositions which are not followed by punctuation marks. As pointed out in 3.2.5, Gries (2002: 232) apparently used such a general search algorithm to identify stranded and pied-piped prepositions in the BNC but obviously failed to extract a substantial number of stranding tokens (Hoffmann 2011: 94). Likewise, syntactic parsing of stranded prepositions is problematic. If a stranded preposition is followed by an adverbial (e.g. *the house which she told you **about** last week*), it is likely that a parser incorrectly analyzes the preposition as belonging to the subsequent adverbial (e.g. *about* might be parsed as belonging to the PP *about last week*). Furthermore, using a fully automatic approach, it would have been impossible to extract tokens with missing prepositions, which are an interesting phenomenon in the context of PCEs (cf. 3.3.3). Consequently, the present study decided against a fully automatic approach.

Now, in contrast to Gries' (2002) automatic procedure, Hoffmann (2011: 18) simply read through the entire ICE-EA corpus. While this method ensures that all stranded and all non-standard preposition placement tokens can be correctly identified, such a time-consuming approach was not feasible in this analysis which includes data from 12 ICE components. The present study thus opted for a semi-automatic method which combines the advantages of automatic extraction and the precision and reliability of a manual approach.

In a first step, the TreeTagger software (Schmid 1994) was used to tag the corpora with POS information according to the BNC Basic Tagset. With the help of an *R* script, the TreeTagger's vertical output was then converted into the horizontal format "word_POS_lemma". Next, another *R* script was written that queried the tagged corpora for all WH-words tagged as WH-determiners, WH-adverbs or WH-pronouns, using the regular expression "\\w+(DTQ|AVQ|PNQ)[^]+". In addition to the WH-word, the *R* script extracted up to 30 elements (i.e. corpus words or ICE markup) standing before and after the WH-word. However, since most text units consist of fewer elements, the majority of the extracted tokens are much shorter. Moreover, the string "<wh>" was inserted before every WH-element to clearly mark the relevant filler (since there are tokens which contain more than one WH-element). In addition, for each token, a version without POS tags was created which was used for the subsequent manual examination of the extracted data. In the same way, all occurrences of *when/where* tagged as subordinating conjunctions were identified (using the regular expression "([Ww][Hh][Ee][Rr][Ee][Ww][Hh][Ee][Nn])_CJS[^]+") as it turned out that the

TreeTagger does not always correctly distinguish between WH-adverbs and subordinating conjunctions.

This procedure yielded a total of 187,158 hits, which were expected to include all WH-tokens licensing variable preposition placement (WH-interrogatives, exclamatives, finite WH-relatives, free relatives, and *it*-clefts; cf. 3.1.1) and all non-standard tokens with a WH-element (doubling, mismatching, missing, resumptive, unexpected, extra tokens; cf. 3.3.3). The only filler-gap constructions which allow for both stranding and pied-piping but could not be extracted are topicalizations (*The English teacher, he is talking to.* – *To the English teacher, he is talking.*) and non-WH-clefts (*It's those patients that I'll be talking about.* – *It's about those patients I'll be talking*). Since both construction types lack a WH-element, these structures could have only been identified by looking for all prepositions preceded or followed by punctuation marks. However, as pointed out above, this procedure is likely to miss a great number of relevant tokens. Consequently, these structures could not be included in the present study.

Now, this corpus query, of course, produced a great number of false positives. Hence, in a second step, a manual examination of all tokens was required. To facilitate this process, all tokens were uploaded to the application *The Red Hen Rapid Annotator* (<https://beta.rapidannotator.org>; as described by Uhrig 2022)¹⁶, which is a user-friendly platform created for the rapid annotation of large amounts of data. One student assistant and I then manually went through all hits to identify all relevant tokens with stranded, pied-piped or non-standard prepositions.

The student assistant was provided with a detailed coding manual and intensive training. There were two annotation levels for each hit: RELEVANCE and PREPOSITION PLACEMENT. At first, the tokens were simply annotated as “relevant” or “NA”. In addition to tokens without stranded, pied-piped or non-standard prepositions, the following cases were coded as “NA”:

- (1) all stranding and pied-piping contexts which do not allow for variable preposition placement, i.e. (cf. chapter 3):
 - a. nonfinite WH-relative clauses:
 - (4.1) *if you don't have a base **from which** to choose* <ICE-CAN:S1B-038#47:1:C>
 - b. finite relative clauses in which the WH-word is extracted out of subject NPs:

¹⁶ I would like to thank Peter Uhrig for his technical support.

- (4.2) *ten to twelve letters to VC **from which** one or two land up on my table for remarks* <ICE-IND:W1B-007#9:1>
- c. tokens in which the WH-element functions as the subject of a passive clause:
- (4.3) *the assumption that **what** is sought **for** must be something definite* <ICE-PHI:W2A-005#68:1>
- (2) in-situ interrogatives consisting of P + FILLER strings only:
- (4.4) *Of what* <ICE-GB:S1A-006#12:1:A>
- (3) hits in which the WH-word belongs to normative insertions, untranscribed text or editorial comments (as e.g. indicated by the tags <&> and </&> in (4.5)), or tokens with a substantial proportion of unclear or missing material (4.6):
- (4.5) *<&> tabulated in the form of each Newspaper and its address **against which** the price is shown </&>* <ICE-IND:W1B-017#15:1>
- (4.6) ***what** are you <unclear> word </unclear> <?> looking **at** </?>* <ICE-CAN:S1B-002#257:2:B>

Moreover, there are tokens such as (4.7) in which speakers repeat the WH-element. Although these utterances surfaced several times in the data (once for each occurrence of the WH-element), they were only included once.

- (4.7) *I don't know like <}> <-> <wh>**what** </-> <=> **what** </=> </}> they cover their teapots with* <ICE-CAN:S1A-032#271:1:A>

All relevant tokens were also annotated for the level PREPOSITION PLACEMENT. This involved the distinction between the categories “stranding”, “pied-piping”, “missing”, “doubling”, “resumptive” and “unsure”, which gave a first impression of the distribution of the data.¹⁷ The “unsure” level allowed the student assistant to quickly deal with ambiguous cases, which were then examined in more detail by myself at the end of the annotation process. As shown in table 4.2, this procedure left a final dataset consisting of 10,079 tokens.

¹⁷ At this stage, no distinction was made between “doubling” and “mismatching” tokens. In addition, in order to annotate the data as quickly and as efficiently as possible, “extra” prepositions, which are sometimes ambiguous at first sight, were simply annotated as “unsure”. A more detailed examination and coding of the non-standard tokens followed later (cf. 4.3.1).

Table 4.2: *Stranded, pied-piped and non-standard tokens across varieties (absolute and relative frequencies)*¹⁸

Variety	Stranding		Pied-piping		Non-standard		Row total	
CanE	438	57.86%	310	40.95%	9	1.19%	757	100%
BrE	452	38.18%	725	61.23%	7	0.59%	1,184	100%
IrE	471	50.27%	446	47.60%	20	2.13%	937	100%
NZE	557	58.63%	382	40.21%	11	1.16%	950	100%
JamE	532	45.78%	601	51.72%	29	2.50%	1,162	100%
SgE	408	54.55%	272	36.36%	68	9.09%	748	100%
KenE	269	36.50%	431	58.48%	37	5.02%	737	100%
TanE	95	18.59%	362	70.84%	54	10.57%	511	100%
HKE	247	40.36%	294	48.04%	71	11.60%	612	100%
IndE	164	17.35%	707	74.81%	74	7.83%	945	100%
NigE	366	46.98%	357	45.83%	56	7.19%	779	100%
PhilE	373	49.27%	352	46.50%	32	4.23%	757	100%
Column total	4,372	43.38%	5,239	51.98%	468	4.64%	10,079	100%

4.3 Data annotation

All relevant tokens were then exported to a spread sheet file in *Libre Office Calc*, where they were annotated for a set of linguistic (cf. 4.3.1) and extralinguistic (cf. 4.3.2) variables.

4.3.1 Linguistic factors

Following previous studies (Gries 2002; Hoffmann 2011; also see sections 3.1 and 3.2), all tokens were coded for the linguistic factors listed in table 4.3. The dependent variable PREPOSITION_PLACEMENT distinguishes between the levels “stranding” and “pied-piping” as well as the non-standard types of preposition placement described in section 3.3.3, i.e. “missing”, “doubling”, “mismatching” and “resumptive” structures. The level “other” accounts for all other cases of non-standard preposition placement such as the token in (4.8), which contains the WH + POSTPOSITION string *what for*:

- (4.8) *Then the question is <,> Mr vice <,> chairman <,,> **what for** <,,> we should be thankful <,,> to the new mens era <ICE-IND:S1B-052#27:1:A>*

¹⁸ In table 4.2, “unexpected” tokens are included as part of the stranding and pied-piping tokens.

Table 4.3: Linguistic factors investigated in the present corpus study

Factors	Levels
PREPOSITION_PLACEMENT	stranding, pied-piping, missing, doubling, mismatching, resumptive, other
NONSTANDARD (non-standard only)	unexpected, extra
CLAUSE_TYPE	exclamative, free RC, embedded interrogative, main clause interrogative, finite RC, cleft RC
PP_TYPE	cf. table 4.4
BE	be, NA
PHRASE_TYPE	VP, AdjP, NP
VOICE	active, passive
RESTRICTIVENESS (RCs only)	restrictive, non-restrictive
COMPLEXITY	number of words between WH-word and gap site
COMPLEXITY_BRIDGE	simple, complex_zero, complex_that
FILLER_LENGTH	length of extracted phrase in number of words
FREQUENCY_FILLER	relative frequencies of fillers computed for each individual ICE corpus
FREQUENCY_PREPOSITION	relative frequencies of prepositions computed for each individual ICE corpus
PREPOSITION	<i>in, for, under</i> , etc.
FILLER_TYPE	<i>who, whom, which, what, whose, when, where, how, wh-ever</i>

In addition, the factor NONSTANDARD was included to account for the fact that “unexpected” (4.9) and “extra” (4.10) prepositions can appear in both stranding and pied-piping constructions.

- (4.9) (a) *What* are you going to concentrate *in* <ICE-PHI:S1A-063#83:1:A>
- (b) *So the development of rural economy depends upon the extent on which the rural development takes place in our country* <ICE-IND:W1A-010#76:1>
- (4.10) (a) *Whom* are you targeting *at* <ICE-EA-T:S1B046T>

Table 4.4: Levels of PP_TYPE (Mußermann 2024: 11; based on Hoffmann 2011: 116)

PP_TYPE	Examples
idiosyncratic	What's it like <ICE-GB:S1A-019 #53:1:B> What did you do that for <ICE-IRE:S1A-053\$A>
VXP	dinner on Friday night which I'm really looking forward to <ICE-GB:W1B-013 #89:2>
prepositional X	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>
obligatory complement	the box which the winch is in <ICE-NZ:S1A-052#163:1:B> the place which he lives in <ICE-GB:S1A-071 #101:1:D>
optional complement	India whose ambassador I talked to <ICE-GB:S2B-010 #49:1:B> anything which I'm interested in <ICE-IRE:S1B-073\$B> the building site which this building has been part of <ICE-IRE:S1B-012\$A>
movement	places where I went to <ICE-IRE:S1A-100\$B>
accompaniment	the people who I work with <ICE-CAN:W1B-020#83:5>
subcategorized PP	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	the paper upon which they are writing <ICE-GB:W2D-006 #40:1>
direction	the direction from which the tram would approach <ICE-CAN:W2F-012#53:1>
location	the room in which they are working <ICE-GB:W2D-006 #93:1>
time	the moment at which Art chose life <ICE-NZ:W2F-006#130:1>
duration/frequency	the time for which each part is flooded <ICE-IRE:W2B-030\$C>
means/instrument	the tools with which they were supposed to be working <ICE-NZ:W2B-009#86:1>
agent	the person by whom he was killed
cause/reason/purpose/result	The reasons for which we neglect taking our medicines <ICE-GB:S2B-038 #72:2:A>
respect	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	the way in which they developed <ICE-GB:S1B-016 #160:1:C>
degree	the extent to which English displaced Latin <ICE-CAN:W2A-008#51:1>

- (b) *we should have some basic information **in which** we will use to question him* <ICE-NIG:parl_09:a122>

The variable `CLAUSE_TYPE` includes six levels (exclamative, free relative (free RC), embedded interrogative, main clause interrogative, finite WH-relative (finite RC), *it*-cleft (cleft RC)), which basically correspond to the construction types discussed in section 3.1.1. An aspect that has not been mentioned yet is the distinction between main clause interrogatives (4.11a) and embedded interrogatives (4.11b). In 3.1.1.2, it was argued that stranding is favored in WH-interrogatives because it ensures the clause-initial placement of the WH-element, which should ideally appear as early as possible for discourse-functional reasons. However, in embedded interrogatives the WH-word is usually preceded by other material regardless of preposition placement. Consequently, the preference for stranding may be weaker in embedded interrogatives than in main clause interrogatives.

- (4.11) (a) ***What** do you want to talk **about*** <ICE-SIN:S1A-033#4:1:B>
 (b) *so you may well wonder **what** i am going to talk **about*** <ICE-NZ:S2A-041#4:1:E>

Furthermore, in line with Hoffmann (2011: 114), tokens such as (4.12a) were coded as pied-piped cleft RCs, while (4.12b) is an example of a stranded *it*-cleft. In contrast, cases such as (4.12c) in which the topicalized element is an NP were coded as pied-piped finite RCs (cf. 3.1.1.1).

- (4.12) (a) *it is to this institution **which** most persons aspire* <ICE-JA:W1A-001#53:3>
 (b) *it is this institution **which** most persons aspire to*
 (c) *it is this institution **to which** most persons aspire*

All other variables (with the exception of `PREPOSITION` and `FILLER_TYPE`) are related to the processing effects outlined in section 3.2. To assess the effect of `PP_TYPE`, the tokens were annotated according to Hoffmann's (2011: 116) fine-grained PP type classification, which distinguishes between a wide range of complement and adjunct PPs (cf. table 4.4). While the PP types at the top of table 4.4, i.e. prepositional verbs or *WH...like* constructions, are assumed to have a very strong stranding preference, the PP types found at the bottom of the table, namely respect, manner and degree PPs, should have an almost categorical pied-piping preference. Moreover, considering that Hoffmann (2011: 139) found that obligatory complements with *be* exclusively co-occur

Another processing-related variable is PHRASE_TYPE, which takes into account that NP-contained PPs are harder to process than AdjP- and VP-embedded PPs (cf. 3.2.3). Furthermore, the factor VOICE, which consists of the levels “active” and “passive”, was included to test Gries’ (2002: 237–238) claim that preposition stranding is avoided with the passive (cf. 3.2.5). The variable RESTRICTIVENESS, which is only relevant for finite RCs, captures the fact that restrictive relatives involve more processing complexity than non-restrictive relative clauses (cf. 3.2.1).

In section 3.2.4, it was argued that the pied-piping probability should increase if the material standing between the filler and the gap is syntactically or semantically complex. However, the semantic complexity of the bridging structure being a complex concept which can hardly be captured by a single variable (Hoffmann 2011: 95–96), in this analysis it was decided to focus on the length of the constructions, which is much easier to measure. While Hoffmann (2011: 97) used an idiosyncratic complexity measurement based on the assumption that a parser always aims to combine individual elements into chunks, in the present study COMPLEXITY was operationalized as the number of words between the WH-word and the gap site (cf. example (4.13a)), which has been shown to be a “near-perfect prox[y]” of syntactic complexity (Szmrecsanyi 2004: 1031). As illustrated in (4.13b), for constructions with pied-piped or missing prepositions, the earliest position in which a stranded preposition can occur was reconstructed (adapted from Hoffmann 2011: 97).

- (4.13) (a) *positive things in there* [which I agree with]
1 2 3 4
<ICE- CAN:S1B-047#40:1:B>
- (b) *a strong leader* with [whom I disagree X]
1 2 3 4
<ICE-CAN:S1B-032#205:2:D>

138

stranded preposition can occur. Then, for each token, all material appearing before the first and after the second tab was deleted so that only the FGD remained.

However, this remaining material still contained the typical ICE markup, which, in most cases, was not supposed to be included in the word count. Hence, using *R*, a list was created which provided an overview of all ICE markers contained in the data. For each marker that surfaced in the list, a decision was made whether it should be deleted or not. While e.g. editorial comments tagged as `<&>...</&>` or elements such as `<}>` or `</}>`, which mark hesitations, had to be removed, in a few cases the markup was relevant for the word count (e.g. the marker `<unclear-word>` represents a spoken word) and thus had to be included. Moreover, spaces before markers such as `<sp>` or `<sb>`, which are used to indicate super- or subscripts, had to be removed to avoid that these elements were counted as individual words.

After deleting all irrelevant markup with the help of *Sublime Text*, I manually went through all tokens from spoken categories. Spoken utterances frequently contain repetitions, self-corrections or disfluencies such as *uh* or *uhm*. As these elements typically only appear in spoken categories, they had to be excluded. Otherwise, tokens from spoken text types such as (4.14), an example from the private dialogue category, would generally tend to be more complex than tokens from written language, where these elements usually do not occur. In contrast, elements which appear to be repetitions but which actually add further information or emphasis, such as *two lines two vectors* in (4.15), were not excluded.

(4.14) *I really wonder to [what extent ~~the-the~~two ~~uh~~ revenue departments ~~uhm~~ ~~uh~~communicate]* <ICE-CAN:S1A-024#97:1:B>

(4.15) *the point at which **two lines two vectors** of trajectory may approach one another* <ICE-CAN:S2B-046#96:1:A>

Moreover, misspellings, such as *foundand* in (4.16), which should be counted as two words, were corrected as it is highly unlikely that the writing as one word reflects an entrenched constructional chunk.

(4.16) *Anyway we had a discussion on the "misterio escondido" of "El celoso extremeno" which I still have not **foundand** think that we should not waste our time looking for* <ICE-JA:W1B-005#57:4>

After these preparatory steps, the final word count was established automatically using an *R* script.

A second complexity-related variable that was coded is the factor COMPLEXITY_BRIDGE, which consists of the three levels “simple”, “complex_zero” and “complex_that”. Following Hoffmann (2011: 193), this variable contrasts “simple” relative clauses such as (4.17a) with more complex constructions that consist of a relative clause which is embedded in another clause. In “complex_zero” constructions (4.17b), this additional clause is not introduced by an overt subordinator, whereas “complex_that” (4.17c) tokens include the complementizer *that* (example (4.17) repeated here from (3.96)).

- (4.17) (a) *I saw the teacher **whom Jane relied on**.*
 (b) *I saw the teacher [**whom you claimed**] Jane relied on.*
 (c) *I saw the teacher [**whom you claimed that**] Jane relied on.*

As outlined in section 3.2.4, in Hoffmann’s (2011: 193) magnitude estimation experiments, BrE speakers rated stranding in simple relative clauses such as (4.17a) better than in syntactically more complex ones such as (4.17b) or (4.17c). The variable COMPLEXITY_BRIDGE thus made it possible to examine how frequently syntactically complex bridging structures such as (4.17b) and (4.17c), in which the filler must be held in working memory over long distances, actually occur in natural language use.

Another factor that is also related to complexity but which has not been mentioned so far is the variable FILLER_LENGTH. While most fillers consist of a single word (i.e. the WH-element), there are also structures such as (4.18) and (4.19) in which the filler is a complex phrase with a nominal head.

- (4.18) *um so there are those people [**whose lives**] we want to find out more about*
 <ICE-NZ:S2A-023#133:1:C>
 (4.19) [**What kind of things**] *were you looking for* <ICE-JA:S1A-042#126:1:A>

While previous research mostly focused on the semantic and syntactic complexity of the bridging structure (cf. 3.2.4), it can be assumed that complex filler phrases such as *What kind of things* in (4.19) also require more semantic processing than simple one-word fillers such as *what* or *which*. Although the length of the extracted phrase did not emerge as a significant variable in Gries’ (2002: 237) analysis, from a processing perspective, it is thus plausible to assume that constructions with long fillers should have a stronger preference for pied-piping, i.e. the variant that is easier to process, than constructions with short fillers. In order to test this claim, the present study included the variable FILLER_LENGTH, which was operationalized as the number of words of the extracted

phrase. However, since filler phrases in pied-piped constructions are always longer than the corresponding stranded variants (due to the presence of the pied-piped preposition), pied-piped prepositions were not counted (compare (4.20a) and (4.20b)).

- (4.20) (a) *[What time] is your exam at tomorrow*
 1 2 <ICE-CAN:S1A-097#101:2:B>
- (b) *Now [at what time] did the game start*
 1 2 <ICE-SIN:S1B-070#285:2:E>

These word counts for FILLER_LENGTH were established using the same semi-automatic procedure as for the variable COMPLEXITY (cf. above).

Furthermore, the variable FREQUENCY_FILLER reflects the degree of entrenchment of the filler (cf. 3.2.5). Gries' (2002: 234) variable FREQ_HEAD simply consisted of the categories "frequent" and "infrequent", but it is unclear how these two levels were defined. As the present study investigates several varieties, it was decided to adopt a more fine-grained approach, and to compute the relative frequencies of fillers for each ICE variety. Note that it was not possible to determine these frequencies using the WH-words that were extracted in order to identify all relevant stranding and pied-piping tokens because these hits also included WH-words belonging to non-corpus material such as editorial comments (cf. example (4.21)).

- (4.21) *This research work was fund aided by <}> <-> Transit </-> <+> Tranzit
 </+> <&> <which is how the company spells it </&> </}> NZ and is cur
 <l^> rently being reviewed prior to full publication <ICE-NZ:W2A-
 040#103:3>*

This was not a problem for the identification of the relevant preposition placement tokens because all hits of the corpus query were checked manually so that stranding and pied-piping constructions that formed part of editorial comments, normative insertions or other irrelevant material could be excluded by hand (cf. 4.2.2). Moreover, for the identification of the relevant stranding and pied-piping tokens, the ICE markup provided helpful information about the correct interpretation of the corpus material. However, for the computation of the filler frequencies, not every WH-token was examined manually. Consequently, all non-corpus material had to be removed from the tagged corpora in order to avoid that irrelevant fillers which are not representative of the corpus language were counted. Just like the removal of the extra-corpus material (cf. 4.2.1), this required the correction of markup errors, which were identified with the help of regular expressions. Then, R scripts were created that queried the cleaned corpora for all WH-words tagged

as WH-determiners, WH-adverbs and WH-pronouns, as well as all instances of *when/were* tagged as subordinating conjunctions (using the regular expressions provided in section 4.2.2). The latter hits were then checked manually in order to identify all tokens in which *when* or *where* is tagged as a subordinating conjunction but actually functions as a filler. Cases such as (4.22) in which *when* is used as a subordinating conjunction were excluded.

(4.22) *After all, I had never been around for him **when** he was growing up* <ICE-SIN:W2F-014#92:1>

Next, the *R* function *unique* was used to get an overview of all filler types contained in the data. This was important to identify misspellings and colloquial forms which had to be corrected or standardized (e.g. *wha* > *what*). Finally, the relative filler frequencies for each ICE corpus were computed with the help of an *R* script.

Besides this, the factor FREQUENCY_PREPOSITION was included in order to shed more light on the claim that more frequent prepositions are more likely to be stranded, which so far lacks a clear explanation (cf. 3.1.2). While Gries' (2002: 234) variable *FREQ-PREP* was based on a frequency ranking of all prepositions, the following hypothetical scenario will show that this approach is not ideal for the present study, which compares preposition placement constructions across several varieties. Let's say the most frequent preposition in variety A accounts for 75% of all prepositions used in variety A. At the same time, in variety B, the most frequent preposition only accounts for 40% of all prepositions occurring in variety B. Nevertheless, in variety-specific frequency rankings, both prepositions would rank first, which would fail to reflect the fact that the most frequent preposition in variety A is used almost twice as often as the most frequent preposition in variety B. Instead of creating a frequency ranking, the present study thus computed the relative frequencies of prepositions for each ICE variety.

These frequencies were determined in the same way as the filler frequencies. The regular expression “`([\\w\\.]+(PRP|PRF)[^ ]+)|(upto_NN1_upto)`” was used to search the tagged versions of the corpora from which all non-corpus material had been removed for all items tagged as prepositions and for all occurrences of *upto*, which is frequently spelled as one word and tagged as a noun. In addition, it was important to take into account that the preposition placement data contained complex prepositions such as *according to* or *due to*, whose relative frequencies also had to be computed. Since the TreeTagger normally only tags the final element of multi-word prepositions as a preposition (i.e. e.g. *due to* is tagged as *due_AJ0_due to_PRP_to*), the complex

prepositions which surfaced in the preposition placement data (*according to, because of, due to, out of, subsequent to*) were identified by means of the following regular expression:

```
"(\\w+_AVP_up (<[>]+> ?)*?<prep>\\w+PRP_to</prep>)|
(according_VBG_[^ ]+ (<[>]+> ?)*?<prep>\\w+PRP_to</prep>)|
(\\w+_CJS_because (<[>]+> ?)*?<prep>\\w+PRF_of</prep>)|
(\\w+_AJ0_due (<[>]+> ?)*?<prep>\\w+PRP_to</prep>)|
(\\w+_AVP_out (<[>]+> ?)*?<prep>\\w+PRF_of</prep>)|
(\\w+_AJ0_subsequent (<[>]+> ?)*?<prep>\\w+PRP_to</prep>)"
```

The relative frequencies of the prepositions were then determined separately for each ICE variety. However, just like the filler data described above, the preposition tokens also contained colloquial forms which had to be standardized (e.g. *bout* > *about* in (4.23)), and irrelevant cases such as (4.24), in which *no* is erroneously tagged as a preposition. These tokens were again identified with the help of the *R* function *unique*.

(4.23) *How' **bout** the <indig> ano </indig> the others when will we buy* <ICE-PHI:S1A-014#36:1:B>

(4.24) *Well_AV0_well </[/> </{/> **no_PRP_no** the_AT0_the thing_NN1_thing is_VBZ_be* <ICE-CAN:S1A-004#241:1:B>

In the final step, an *R* script was created which computed the relative frequencies of prepositions for each ICE corpus.

The remaining two variables capture idiosyncratic lexical effects. While PREPOSITION was coded to assess the preposition placement preferences of individual prepositions, FILLER_TYPE accounts for idiosyncratic effects of individual fillers (in particular *who* and *whom*; cf. 3.1.2). In order to identify the prepositions, for stranded, doubled and resumptive tokens, I relied on the version of the data which only included the FGDs (which were extracted in order to perform the word count for COMPLEXITY; cf. above) and extracted the final word of each token. For pied-piped prepositions, I used the tagged version of the data and extracted the lemma of the preposition appearing before the filler (using the regular expression "(?<=_PR[PF]_)[^]+(?<= (<[>]+>)*)?<wh>)"). Then, all extracted prepositions were additionally checked manually in order to correct cases in which the relevant preposition could not be identified because it was erroneously tagged as an adverb, and to account for multi-word prepositions such as *due to* or *because of*, in which only the final element is tagged as a preposition. In order to identify the

fillers, I also used the version of the data which only included the FGDs (without pied-piped prepositions) and extracted the first word of each token. Moreover, all colloquial forms of prepositions and fillers were standardized.

Since the data annotation required a lot of manual coding, the study relied on the support of five student assistants. The variables `CLAUSE_TYPE`, `RESTRICTIVENESS`, `COMPLEXITY_BRIDGE`, `PHRASE_TYPE` and `VOICE` were coded by two students each. All coders received detailed instructions and intensive training. Furthermore, in order to assess the degree of agreement among the coders, each coder annotated a training sample. The size of the training sample depended on the variable that was coded (cf. table 4.5). Variables which consist of a larger number of levels (e.g. `CLAUSE_TYPE`) require a smaller sample size than variables with fewer levels such as `VOICE` or `RESTRICTIVENESS` (Bujang & Baharum 2017: 6). The largest training samples were coded for `PHRASE_TYPE` and `COMPLEXITY_BRIDGE` because the data contained only few observations for the NP-level of `PHRASE_TYPE` and the `complex_zero/complex_that`-levels of `COMPLEXITY_BRIDGE` (cf. 5.2).

Table 4.5: Interrater agreement: sample size and Kappa scores

Variable	Sample size	Cohen's kappa
<code>PHRASE_TYPE</code>	450	0.91
<code>CLAUSE_TYPE</code>	150	0.96
<code>COMPLEXITY_BRIDGE</code>	450	1
<code>RESTRICTIVENESS</code>	200	0.81
<code>VOICE</code>	200	0.92

The comparability of these codings was then evaluated using Cohen's kappa, a measure defined as “the proportion of agreement beyond that expected by chance, that is, the *achieved* beyond-chance agreement as a proportion of the *possible* beyond-chance agreement” (Sim & Wright 2005: 258; original emphasis). As shown in table 4.5, the kappa scores for all variables are ≥ 0.81 , which suggests that the interrater agreement was “almost perfect” (Landis & Koch 1977: 165).

4.3.2 *Extralinguistic factors*

The extralinguistic variables that were coded are listed in table 4.6. While the factor `VARIETY` accounts for variety-specific preposition placement preferences, `VARIETY_TYPE` aims to detect effects related to the Dynamic Model.

Table 4.6: *Extralinguistic factors investigated in the present corpus study*

Factors	Levels
VARIETY	BrE, CanE, IrE, NZE, JamE, SgE, HKE, KenE, TanE, IndE, NigE, PhilE
VARIETY_TYPE	III, IV, V
CATEGORY	private dialogue, public dialogue, unscripted monologue, scripted monologue, broadcast news, written as spoken, non-professional writing, private correspondence, business correspondence, legal presentations, academic writing, non-academic writing, reportage, instructional writing, persuasive writing, creative writing
TOPIC	administration, art and literature, Asia, court, disease, economy, education, government, history and culture, household and gardening, language, national development, news, personal, philosophy, science, situational descriptions, society, sports, technology

Furthermore, recall that Standard Englishes favor pied-piping in formal registers, whereas stranding is preferred in informal text types (cf. 3.1.1.1). This effect is captured by two factors, viz. CATEGORY and TOPIC. The levels of CATEGORY correspond to the ICE text types. Note that the levels “written as spoken” and “legal presentations” only apply to ICE-EA. In addition, the factor TOPIC was included as “a more fine-grained version of the variable FORMALITY” (Funke & Bernaisch 2022: 10). In section 4.1, it was pointed out that the individual ICE components are not always perfectly comparable because not all text types could be collected in all varieties (cf. 4.1). Moreover, it can be expected that the corpus texts differ in terms of their content. For instance, Degenhardt & Bernaisch (2022: 211, fn.4) suggested that “TOPIC captures the content of speaker texts more readily than the rather language-external ICE-based genre distinctions”. In order to account for these effects, the present study included the variable TOPIC (also see Mußemann 2024).

To assign topics to the corpus files, a topic model was fitted with the help of the *R* package *topicmodels* (Grün & Hornik 2022). Since many ICE files consist of several texts, in a first step, all ICE files were split into subtexts to avoid topic assignments that

are very broad (see Steyvers & Griffiths 2007: 441).¹⁹ Next, using the *R* package *tm* (Feinerer & Hornik 2020), all corpus words were stemmed, and punctuation marks, numbers and so-called “stopwords”²⁰, i.e. semantically insignificant words such as *did*, *which* or *each*, were removed from the data as they are not relevant for the generation of topics. Moreover, once again, irrelevant ICE markup had to be deleted. Then, a document-term matrix was created which described the frequencies of all terms occurring in the data. This document-term matrix served as the basis for the topic model (also see Grün & Hornik 2011: 12).

The *R* package *topicmodels* can be used to fit topic models based on Latent Dirichlet Allocation (LDA). According to Blei & Lafferty (2009: 73),

[t]he idea behind LDA is to model documents as arising from multiple topics, where a *topic* is defined to be a distribution over a fixed vocabulary of terms. Specifically, we assume that K topics are associated with a collection, and that each document exhibits these topics with different proportions. This is often a natural assumption to make because documents in a corpus tend to be heterogeneous, combining a subset of main ideas or themes that permeate the collection as a whole. (Blei & Lafferty 2009: 73; original emphasis)

After the user has specified the number of topics to be modeled, LDA assigns each word in the data randomly to one of the topics. These topic assignments are then iteratively updated, taking into account the topic assignments of all other words (Steyvers & Griffiths 2007: 436). Finally, two *R* functions can be used to obtain the most likely topics for each text file as well as the most likely terms for each topic (Grün & Hornik 2011: 11). This allows the development of topic labels and the assignment of a topic to each text file. The output that can be obtained in this way is illustrated in table 4.7, which gives the most prominent terms of three topics that were labeled “economy”, “government” and “technology” (for the most likely terms of all topics, see table A.1 in the Appendix). An overview of all topics included in this study is provided in table 4.6.

What should be emphasized is that, although topic models are “a relatively objective data-driven technique” (Murakami et al. 2017: 272), it is the user who chooses the number of topics. As Murakami et al. (2017: 250) pointed out,

[t]here is no agreed way to automatically decide the number of topics [...]. In other words, the decision on how many topics a corpus will be deemed to contain is a subjective one and the answer may be defended on the grounds of usefulness but not on the grounds of accuracy.

In the present study, the results of several topic models were compared and the optimal number of topics was determined based on a manual evaluation process. In the end, a

¹⁹ Exceptions are ICE-NIG, which only includes one text per file, and ICE-EA and ICE-IRE because in these corpora the different design of the text unit markers does not always allow the assignment of individual extracted tokens to subtexts.

²⁰ For an overview, see <http://snowball.tartarus.org/algorithms/english/stop.txt> (25 January, 2022).

topic model with 20 topics was accepted as the best result. Although this model produced some semantically related topics (such as economy and national development), it yielded better results than models with fewer topics which lacked potentially important topics or did not distinguish well between individual text files. At the same time, topics obtained from models with more than 20 topics turned out to be semantically incoherent and could not be clearly distinguished from each other.

Table 4.7: The 20 most likely terms for the topics economy, government and technology

TOPIC	economy	government	technology
most likely terms	year	govern	system
	compani	parti	comput
	market	state	design
	busi	member	oper
	dollar	elect	inform
	bank	polit	program
	price	presid	model
	pay	minist	process
	cost	nation	data
	rate	speaker	set
	employ	public	problem
	money	issu	work
	cent	law	requir
	increas	committe	technolog
	million	vote	machin
	govern	hous	applic
	tax	nigeria	servic
	hundr	council	network
	percent	countri	power
	thousand	constitut	product

4.4 Statistical analysis

The present study uses three different statistical techniques to analyze the preposition placement tokens extracted from the corpora: mixed-effects logistic regression modeling, Hierarchical Configural Frequency Analysis (HCFA), and “Large Number of Rare Events” (LNRE) models. In the following sections, these methods will be explained in more detail.

4.4.1 *Mixed-effects logistic regression modeling*

In order to investigate how processing constraints and stylistic variables affect the competition between preposition stranding and pied-piping in different varieties, the study draws on mixed-effects logistic regression modeling (Gelman & Hill 2007; Baayen 2008: 241–302), which is the ideal type of regression modeling to be used when dealing with a categorical response variable such as PREPOSITION_PLACEMENT (see Winter 2020: 198). Mixed-effects logistic regression analysis has the advantage that it does not only take into account the effects of “fixed” variables, i.e. constant factors which “have a predictable non-idiosyncratic influence on the response that could be tested with new samples” (Winter 2020: 236), but also so-called “random” effects. According to Gries (2021: 771), a variable qualifies as a random effect

if the levels of the variable in the sample do not exhaust the levels that the variable would have in the population. [...]

if the real interest is not so much in what exactly these particular speakers and stimuli do but in generalizing from them to the population and just controlling for speakers’ idiosyncrasies and for those of the stimuli.

In the context of corpus linguistics, the modeling of random effects is thus based on the assumption that model variables affect individual corpus speakers and individual linguistic items that surface in the data in different ways. The estimated random effect variance in a mixed model can be said to reflect the average adjustments required to account for differences between individual speakers or individual items. According to Baayen (2008: 242),

[t]he adjustments required for individual subjects [or items] will in general not be zero, instead, they will vary around zero with some unknown standard deviation. In mixed models, the standard deviations associated with random effects are parameters that are estimated, just as the coefficients for the fixed effects are parameters that are estimated.

Importantly, though, as noted by Winter (2020: 238), “mixed models do not actually estimate one parameter per participant [or item] in [a] dataset. Instead, mixed models estimate the *variation around the specified random effect*”.

With respect to the present study, this suggests that there are two variables which should definitely be considered random effects: PREPOSITION and FILE_ID. That individual prepositions have idiosyncratic preposition placement preferences has already been pointed out in section 3.1.2. Moreover, the stranded and pied-piped prepositions included in the present study most likely do not cover all preposition types that exist in English. For instance, if the present study was repeated using a different database, the results could be expected to differ from those obtained in the present analysis as a different corpus would not contain exactly the same prepositions as the ICE data. Hence,

it will only be possible to generalize from the ICE corpora to the population if PREPOSITION is included as a random effect.

But why is it that PREPOSITION should be treated as a random effect but not FILLER_TYPE? Recall that, for example, *what*, *how* or *wh-ever* elements are also very strongly associated with specific types of preposition placement (cf. 3.1.2). However, in contrast to the idiosyncratic effects of individual prepositions, the preposition placement preferences of individual fillers can be explained by a correlation between filler type and clause type (the only exception being the effect of *who/whom*). As Hoffmann (2011: 41) pointed out, “[t]he type of displaced element does not influence preposition placement. Instead, the type of displaced element is the result of the specific clause type that it occurs in”. Hence, the effect of FILLER_TYPE can be captured by the variable CLAUSE_TYPE and does not have to be included as a random effect (also see 5.2.2.1).

The second variable that qualifies as a random effect in the present study is FILE_ID. Just like individual prepositions have idiosyncratic preposition placement preferences, individual speakers can also be expected to vary in their baseline preference for preposition stranding and pied-piping. In other words, all data points produced by a single speaker can be assumed to be “more similar to each other than data points from different [speakers]” (Winter 2020: 232). If these dependencies in the data are not accounted for, the risk of obtaining false positives (i.e. the Type I error rate) increases (Winter 2020: 232–233). Moreover, the ICE corpus speakers represent a small sample of the population. Consequently, the results of a regression model based on the ICE data can only be generalized to the population if the model estimates a parameter for the by-speaker variation. The present study will thus include the variable FILE_ID as a proxy for speaker identity.²¹

Now, so far, it has only been mentioned that the overall preference for preposition stranding and pied-piping may vary across prepositions and speakers. This effect can be accounted for by fitting random **intercepts**, i.e. by including “a standard deviation that represents the by-participant [or by-item] variation around the overall intercept” (Winter 2020: 238). However, speakers and prepositions may not only differ in their general tendency to choose stranding or pied-piping, i.e. with respect to the “dependent” variable, but it might also be the case that they interact in idiosyncratic ways with other predictors (Gries 2021: 758), i.e. with so-called “independent” variables (for an overview of

²¹ Although most ICE-components include speaker IDs, the variable FILE_ID is used because most corpus speakers contribute only a single preposition placement token.

dependent and independent variables, see e.g. Gries 2009: 12). For example, speakers may respond differently to a level change of the predictor `CLAUSE_TYPE`. While speaker A may strongly favor pied-piping in finite RCs and stranding in free RCs, speaker B may always favor pied-piping regardless of the clause type. This type of by-speaker variation could be captured by adding another variance term, namely by fitting a by-participant **slope** for `CLAUSE_TYPE` (see Winter 2020: 238). In the same way, additional variance terms could be included for all variables of a model.

Note, however, that there is no general consensus on the type of model selection approach which should be used to identify the optimal random effects structure. On the one hand, there is the so-called “keep it maximal” approach (Barr et al. 2013), which claims that models should always include the maximally possible random effects structure. Arguing that “there are few circumstances in which we can know *a priori* that a random-intercepts-only model is truly justified” (Barr et al. 2013: 274), advocates of the maximal approach recommend that, “for whatever fixed effects are of critical interest, the corresponding random effects should be present in that analysis” (Barr et al. 2013: 276). Put differently, a maximal model should account for all cases of by-subject or by-item variation by estimating variance terms for random intercepts, random slopes, and all correlations between random effects.

On the other hand, such a maximal approach frequently leads to practical problems and cannot be implemented due to convergence issue (Bates, Kliegl, et al. 2015; Brauer & Curtin 2018: 400). Furthermore, although maximal models reduce the Type I error rate, they tend to have lower statistical power, which makes it harder to detect significant fixed effects. Consequently, these models are likely to suffer from Type II errors (Matuschek et al. 2017: 306; Seedorff, Oleson & McMurray 2019). Criticizing Barr et al.’s (2013) maximal approach, Matuschek et al. (2017: 312) thus proposed that the goal of model selection should always be “to identify the most parsimonious model that can be assumed to have generated the data”. In a similar vein, Bates, Kliegl, et al. (2015) pointed out that it can be reasonable to exclude variance terms and correlation parameters from a model if the researcher’s primary focus is on the fixed effects:

[W]e assume that the researcher has hypotheses about main effects and (some) interactions, but that he/she has no specific expectations about variance components or correlation parameters. In other words, we assume that the experimental hypotheses relate to the fixed effects, not the random-effects structure. In hypothesis testing, usually the primary reason for dealing with the random-effects structure is to obtain as powerful tests as justified of the fixed effects. Therefore, it is reasonable to remove variance components/correlation parameters from the model if they are not supported by the data. If there are specific expectations about, say, a correlation parameter, it makes sense to include it in the model (as well as the related variance components). (Bates, Kliegl, et al. 2015: 4)

Parsimonious models do not just reduce the risk of Type II errors, but they also avoid overfitting (Matuschek et al. 2017: 308), which is a great risk in studies such as the present one which deal with low token numbers of individual speakers. Instead of fitting maximal models that estimate parameters for all possible variance terms, it is thus frequently more practical to only include those random effects for which a real theoretical motivation exists (Winter 2020: 242).

In this analysis, generalized linear mixed-effects models will be fitted with the help of the *R* package *lme4* (Bates, Mächler, et al. 2015). The study will apply treatment coding, which means that one level of each categorical variable serves as a reference level to which all other levels of the predictor are compared. In the light of the controversy over model selection described above, the following backward elimination procedure will be used to identify the final structure of a model (see Zuur et al. 2009: 120–123; Gries 2021): With respect to the fixed effects, a maximal model will contain all theoretically motivated predictors as well as all interactions for which both a linguistic motivation and sufficient data exist. Regarding the random effects, random slopes and correlation parameters will only be estimated if there is a reasonable motivation for including them and if the token numbers are sufficiently high to avoid convergence issues and overfitting. In a top-down approach, such a maximal model will then be simplified based on the results of likelihood ratio tests (LRTs) performed with the help of the *R* functions *anova()* and *drop1()*. In this way, first non-significant random effects and then non-significant fixed effects will be removed until the final model only contains variables which are supported by LRTs (using the significance threshold $p < 0.05$).

After determining the final model structure, a set of diagnostic tools will be used to evaluate the validity of the model. First of all, it is important to ensure that the model predictors do not exhibit multicollinearity. According to York (2012: 1380), collinearity can be defined as a situation “when X_1 is sufficiently highly correlated with another factor, X_2 , so that it is difficult or impossible to disentangle the independent effects of X_1 and X_2 ”. Consequently, the estimates of a model in which two or more predictors are strongly correlated are unreliable, which means that the risk of obtaining a Type II error increases (York 2012: 1384). In the present study, two parameters will be used to assess multicollinearity: variance inflation factors (VIFs) and the condition index K . The VIF is a measure which shows “how much the variance of the estimate of the coefficient of an independent variable is inflated compared to a case of no collinearity” (York 2012: 1382). According to Montgomery & Peck (1992), VIFs should remain below 10, while Zuur, Ieno & Elphick (2010: 9) suggest that the VIFs of all predictors should be as low as 3.

Importantly, however, the VIF depends on the contrast-coding scheme used for a factor (Fox & Weisberg 2019: 432; for an overview of contrast codings, see e.g. Winter 2020: 117–132). Fox & Monette (1992) thus developed an alternative measure called “generalized variance inflation factor” (GVIF) which is independent of the contrast coding of a variable. The present study will use adjusted GVIFs ($\text{GVIF}^{1/(2 \cdot \text{DF})}$; with DF being the degrees of freedom of the predictor variables) as implemented in the *R* package *car* (Fox & Weisberg 2019) because this measure allows comparisons between predictors with more than two levels (Fox & Monette 1992). The second measure used to assess collinearity is the condition number K (Belsley, Kuh & Welsch 1980). According to Baayen (2008: 182), “[t]he condition number estimates the extent to which a matrix is singular, i.e. how close the task of estimating the parameters is to being unsolvable”. While condition numbers between 0 and 6 suggest that there is no collinearity, “potentially harmful collinearity” is indicated by numbers above 30 (Baayen 2008: 182).

Furthermore, in order to evaluate the overall significance of the final model, a null model will be fitted which does not contain any fixed effects but the same random effects terms as the final model. Then, an LRT between the two models will be performed to test if the final model is significantly better than the null model (Gries 2021: 776–777). Besides this, the study will compute so-called Pseudo- R^2 s (Nakagawa & Schielzeth 2013). Pseudo- R^2 s, which have values reaching from 0 to 1, contain information about the proportion of variance explained by a model. Compared to information criteria such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), Pseudo- R^2 s have the advantage that they can be used for comparisons across different datasets (Nakagawa & Schielzeth 2013: 134). For mixed-effects models, there are two types of Pseudo- R^2 s: While the marginal R^2 provides information about the amount of variance explained by the fixed effects, the conditional R^2 indicates how much variance is accounted for by both the fixed and the random effects (Nakagawa & Schielzeth 2013: 136). The difference between the two values indicates the amount of variance explained by the random effects (Nakagawa & Schielzeth 2013: 140).

In addition, the classification accuracy of the final model will be determined with the help of the *R* package *caret* (Kuhn 2008). The classification accuracy describes the proportion of correct predictions out of the total number of predictions made by a model (Gries 2021: 777). However, as the following example from Bernaisch (2022: 176) shows, this score must be compared to the accuracy score of a model that always chooses the most frequent option:

For linguistic objects of investigation, a [model] with a classification accuracy of 80.15% at first sight appears to do a great job at predicting [...] – it predicts more than four out five choices

correctly. Still, if we had a much simpler statistical model (i.e. one that always predicts the more/most frequent choice in the observed values [...] all the time), we would already get a classification accuracy of 76.6%. (Bernaisch 2022: 176)

Moreover, it is useful to complement accuracy values with precision and recall scores (Gries 2021: 768). In the present study, precision describes the proportion of pied-piping predictions that are actually correct, while recall provides information about the proportion of actual pied-piping cases that are predicted correctly (Hahn 2005: 534). Another common measure used to evaluate model predictions is the *C*-score, which can be described as “an index of concordance between the predicted probability and the observed response” (Baayen 2008: 182). The values of *C* range from 0 to 1, with 0 indicating random prediction and 1 indicating perfect prediction. According to Baayen (2008: 182), “[a] value above 0.8 indicates that the model may have some real predictive capacity”.

4.4.2 Hierarchical Configural Frequency Analysis (HCFA)

The second statistical method used in the present study is Hierarchical Configural Frequency Analysis (HCFA). HCFAs will be performed with the help of Gries’ (2004) *HCFA 3.2* script for *R*. HCFAs test whether combinations of two or more factor levels, i.e. so-called “configurations” (Gries 2009: 244), occur more frequently or less frequently than expected by chance. Configurations which are observed more frequently than expected by chance are so-called “types”, while configurations which occur less frequently than expected are considered “antitypes” (Von Eye 2002: 3; Gries 2009: 244). An HCFA subjects all configurations to a chi-square test and assesses if the frequency deviations are significant. When the expected frequencies of a configuration are lower than 5, the *HCFA 3.2* script uses exact binomial tests (Hoffmann 2011: 26).

The output of Gries’ (2004) *HCFA 3.2* script comes in the form of a table which lists all configurations that were tested. As shown in table 4.8, for each configuration, the table provides the observed and expected frequencies, a greater-than or less-than sign which illustrates whether a configuration is a type or an antitype, the configuration’s contribution to chi-square, a *p*-value adjusted for multiple *post hoc* testing, a series of asterisks illustrating the significance value (* for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$) and a “coefficient of pronouncedness” *Q*, which provides information about the effect size, with larger *Q* values indicating stronger configurations (Gries 2009: 251).

Since the token numbers in the present analysis vary considerably across varieties (cf. table 4.2), for this study, it is particularly important that HCFAs can be used to analyze data which are distributed unevenly (Hoffmann 2019: 206). At the same time, it

must be kept in mind that HCFAs are conservative and unlikely to detect significant effects with small sample sizes, which increases the risk of Type II errors (Hoffmann 2021: 32). Consequently, in the present analysis, HCFA results that are based on low token numbers will have to be taken with a grain of salt.

*Table 4.8: Example: HCFA output for VARIETY_TYPE * PREPOSITION * FILLER*

VARIETY_TYPE	PREPOSITION	FILLER	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
III	to	what_extent	37	9.4183	80.7736	>	2.137703684878E-08	***	0.073
III	for	how_long	22	3.1044	115.0121	>	1.341630737567E-08	***	0.049
IV	for	how_long	8	0.7319	72.1755	>	0.006901220189698	**	0.019
V	on	what_basis	6	0.2128	157.3857	>	0.000711548800805	***	0.015
V	at	what_point	6	0.2549	129.4868	>	0.002029187018272	**	0.015

In this investigation, HCFAs will be used for two different types of analysis. In section 4.4.1, it was pointed out that mixed-effects models can only be used to test interactions of predictors if there is sufficient data. This is of major importance to avoid overfitting, i.e. “situations when models correspond too closely to the idiosyncratic patterns of particular datasets” (Winter 2020: 115). In a few cases in which the token numbers are not sufficient for the fitting of a generalized linear mixed-effects model, HCFAs will thus be used to assess if certain combinations of factor levels occur more or less frequently than expected by chance. As it is frequently impossible to include three-way or four-way interactions of categorical predictors in logistic regression models, HCFAs will be particularly important to test the significance of configurations involving three or more categorical variables. However, as HCFAs are unable to account for by-item and by-subject variation, whenever possible, of course, mixed-effects models will be preferred.

A second aspect for which HCFA will be applied has to do with Hoffmann’s (2021) Dynamic Model Productivity Hypothesis (cf. 2.2.3.2). Since HCFAs test whether certain combinations of two or more factor levels are particularly frequent or infrequent, they are not only useful to reveal whether any combinations of variables such as CLAUSE_TYPE or PHRASE_TYPE occur more frequently than expected by chance, but they can also identify significant associations between fillers of constructional slots and specific varieties or variety stages (Hoffmann 2021: 29). Following Brunner &

Hoffmann (2020), HCFA will thus also be applied to identify entrenched meso-constructions (also see Brunner 2021). This approach is illustrated in table 4.8, which shows the output of an HCFA which tested associations between VARIETY_TYPE, the PREPOSITION slot, and the FILLER slot in pied-piped interrogative constructions. In line with Brunner & Hoffmann (2020: 21), each significant P + FILLER combination that is associated with a variety stage can be said to “represent[...] a separate node in the respective speakers’ constructional network”²². Consequently, HCFA offers a useful statistical tool to investigate the claim that “varieties at earlier stages of the Dynamic Model show a greater reliance on fewer and more prototypical and frequent fillers than varieties at later stages” (Hoffmann 2021: 29).

4.4.3 *Large Number of Rare Events (LNRE) models*

While HCFA will be used for a qualitative analysis of the meso-constructions that are associated with individual varieties or variety types, the present investigation also aims to test another prediction made by Hoffmann’s (2021) Dynamic Model Productivity hypothesis (cf. 2.2.3.2), namely the claim that “[v]arieties in later phases of the Dynamic Model [...] show more productivity of the slots of (semi-)schematic construction than varieties in earlier developmental phases” (Hoffmann 2021: 21). Following previous studies (Brunner & Hoffmann 2020; Hoffmann 2020a, 2021; Brunner 2021), the study will use so-called “Large Number of Rare Events” (LNRE) models (Evert 2004; Baayen 2008: 229–236; Zeldes 2013; Baroni & Evert 2014) to compare the productivity of constructional slots across variety stages. Compared to other measures of productivity, such as type token ratios or the ratio of hapax legomena, i.e. of types with a frequency of 1 (see Baayen 2008: 223), LNRE models have the advantage that they take into account that the distribution of linguistic elements corresponds to Zipf’s law, which means that there is a small number of words with a very high frequency, while most linguistic items are “low probability elements” (Baayen 2008: 229) that are unlikely to surface in a given sample (Evert 2004; Hoffmann 2021: 28). This Zipfian distribution is particularly problematic in studies such as the present one which aim to compare the productivity of

²² Remember, however, that such HCFA results are based on aggregated data from several speakers. As pointed out in section 1.2.2, significant associations between fillers of constructional slots and varieties/variety types should thus not be assumed to reflect direct cognitive entrenchment. Instead, configurations that emerge as significant in an HCFA can be said to indicate conventionalized meso-constructions of the respective varieties/variety types (see Schmid 2020: 270).

constructional slots across corpora that differ in size (recall that ICE-EA-T and ICE-EA-K are smaller than the other ICE components; cf. 4.1). While small corpora can be expected to contain a large number of new types, the type frequency should decrease considerably with increasing sample size. Consequently, traditional measures such as type-token ratios, which fail to take into consideration this Zipfian distribution, will yield inaccurate estimates which cannot be compared across varieties (Baayen 2008: 229; Hoffmann 2021: 28).

The present study will therefore assess the productivity of constructional slots using finite Zipf-Mandelbrot models²³ (Evert 2004) fitted with the help of the *R* package *zipfR* (Evert & Baroni 2007). As Hoffmann (2021: 29) pointed out, the great advantage of the *zipfR* package is that it

tracks the number of new types as the sample size increases (its ‘vocabulary growth’). It then does not project the productivity of a pattern in a linear fashion (since new types become rarer and rarer as sample size increases). Instead, LNRE models project the productivity of a schematic slot beyond the empirical sample size by predicting a curve that first rises steeply and then seems to flatten (since it is a type of cumulative distribution function of an exponential type of distribution).

The *zipfR* package thus makes it possible to extrapolate productivity curves from the observed data up to a sample size specified by the user. The output that can be obtained in this way is a plot that includes productivity curves and confidence intervals for each variety type. Moreover, vertical lines can be added to indicate the actual token numbers in the corpora. These plots can be used for a reliable comparison of the productivity of constructional slots across variety stages. If the confidence intervals of growth curves from different variety phases overlap, this can be taken as an indication that the productivity rate does not differ significantly between variety stages (Brunner 2021: 24). In addition, the goodness of fit of an LNRE model can be evaluated based on the results of a multivariate chi-squared test that comes with the summary of the model parameters. According to Baroni & Evert (2014: 10), a good model fit is reflected by a low chi-squared statistic and high *p*-values. Baayen (2008: 233) suggests that *p*-values should ideally be greater than 0.05.

This section outlined the methodology used in the present study. While 4.1 introduced the ICE project, 4.2 and 4.3 described how the data were extracted and annotated. Section 4.4 focused on the three different statistical techniques that are relevant for the corpus

²³ The *zipfR* package also supports Zipf-Mandelbrot models. However, as Evert (2004: 7) pointed out, “its assumption of an infinite vocabulary is unrealistic for natural-language data”. Hence, the study will use finite Zipf-Mandelbrot models.

analysis. In the next chapter, these statistical methods will be applied to investigate the constructional competition between preposition stranding and pied-piping in World Englishes.

5. Preposition placement in World Englishes

This chapter provides a corpus-based study of preposition placement in World Englishes. While section 5.1 will outline the hypotheses to be tested, the subsequent sections (5.2 – 5.5) will give the study results. 5.6 will provide a final discussion and summary of all findings.

5.1 Hypotheses

In the previous chapters, it was shown that the competition between preposition stranding and pied-piping is largely driven by processing factors and formality effects (cf. 3). In line with Hawkins' (2004) PGCH, I expect that speakers of all varieties should have a preference for preposition placement constructions that are easy to process (cf. 3.2). At the same time, following Hoffmann (2011), I predict that speakers of PCEs, who can be assumed to receive less input, have a greater preference for prototypical constructions than L1 speakers. Moreover, I expect that the limited amount of input in PCEs leads to a lower degree of stylistic variation in these varieties (cf. 2.3). These assumptions can be summarized in the following hypotheses:

Table 5.1: Prototypical stranding and pied-piping constructions (repeated from 3.2.6)

Prototypical stranding constructions	Prototypical pied-piping constructions
	Pied-piped (restrictive) finite WH-relative clause construction
Stranded constructions with lexicalized verb-preposition strings	Pied-piped constructions with time, location, direction, instrument, contingency, frequency, respect, manner, and degree PPs
	Pied-piped constructions with NP-embedded PPs
	Pied-piped passive constructions

- 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; cf. table 3.2 repeated here as table 5.1).

- 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. While Phase-V varieties should have a clear preference for pied-piping in formal registers and for stranding in informal contexts, varieties at earlier stages of evolution should always prefer the more prototypical and formal variant, i.e. pied-piping.

Furthermore, in line with Hoffmann's (2021) Dynamic Model Productivity hypothesis, I expect that varieties at higher developmental stages exhibit more variation in constructional slots than varieties at lower Dynamic model stages. Consequently, Phase-V varieties should show a high type frequency of meso- and macro-constructions, whereas varieties at lower stages of evolution should rely on a small set of frequent and prototypical meso-constructions (cf. 2.2.3). This claim is reflected in hypotheses IV and V:

- IV. There should be more entrenched meso-constructions (i.e. significant associations between fillers of constructional slots and specific varieties) in more advanced varieties than in less advanced varieties. At the same time, the slot fillers identified as being significantly associated with varieties at lower stages of evolution should be prototypical and frequent elements (e.g. P + WH combinations such as *in which* or *to whom*, or V + P combinations such as *rely on* or *talk about*).
- V. The productivity of preposition placement constructions should correlate with Dynamic Model phases, with Stage-III varieties displaying the lowest and Stage-V varieties displaying the highest slot productivity.

In addition, following Hoffmann (2021), I assume that second-language learning principles and processing factors lead to qualitative innovations in PCEs which should emerge at the meso-constructional level (cf. 2.2.3.1). This prediction is captured by hypothesis VI:

- VI. PCEs should be characterized by innovative uses of partly schematic preposition placement constructions, such as structures with doubled, missing, unexpected or extra prepositions or constructions with resumptive pronouns.

These innovations should be motivated by processing factors and general second-language learning principles.

Moreover, I expect that the constructional choices of speakers of PCEs are affected by the preposition placement constructions that exist in the local languages of these varieties (cf. 2.4.3). This claim is summarized in hypothesis VII:

- VII. PCEs should exhibit quantitative innovations which can be attributed to cross-linguistic influence. In accordance with the claims made in sections 3.4.1 and 3.4.2, speakers of varieties with local languages in which pied-piping constructions are the norm should have a preference for pied-piping in English. Speakers of varieties with local languages in which preposition stranding dominates should favor stranding in English.

5.2 Preposition stranding and pied-piping in World Englishes

Hypotheses I, II, III and VII were tested with the help of generalized linear mixed-effects modeling. Since the study aims to detect variety-specific patterns as well as effects related to the Dynamic Model, separate models had to be fitted for the variables VARIETY_TYPE (cf. 5.2.2) and VARIETY (cf. 5.2.3) to avoid problems with rank deficiency. In order to take a closer look at preposition placement in relative clauses, i.e. the most complex construction type included in the present analysis (cf. 3.2.1), an additional model was fitted based on the relative clause data only (cf. 5.2.4). However, before the results of these models will be presented, a set of tokens will be described which had to be excluded from the logistic regression analysis.

5.2.1 Excluded tokens

There were four groups of tokens which had to be eliminated from the data before model selection could begin. First of all, the data contained 14 tokens for which not all variables could be coded. For instance, in (5.1), it is unclear whether CLAUSE_TYPE is a finite RC, an embedded interrogative or a free RC construction. In (5.2), the absence of the verb makes it impossible to decide whether the token should be coded as an obligatory complement PP (i.e. *where they are from*), as a movement PP (i.e. *where they come from*) or as a different type of PP. Since the *lme4* package removes data with missing values, these tokens could not be included in the logistic regression analysis.

- (5.1) *Though he's gone <,> and these <,> given ad in the paper and all that <,> its for sale you know <,> and this <O> two words </O> **from where this ad came** <ICE-IND:S1A-041#100:1:A>*
- (5.2) *Over here we have two other bird <unclear> words </unclear> teach you **where they from** <ICE-JA:S2A-053#X159:1:F>*

Moreover, in the entire dataset, there were only 5 exclamative and 7 cleft RC tokens. As expected, all of the exclamative tokens have stranded prepositions (5.3a-e), which supports the claim that exclamative constructions favor stranding due to the discourse function of the WH-word, which signals exclamation and should thus appear as early as possible. Interestingly, three of these tokens (5.3a-c) involve the pattern *what...BE...in*. The investigation of a partly schematic *what...BE...in*-construction might thus be an interesting aspect for future research based on larger corpora.

- (5.3) (a) ***What** a mess she was **in** <ICE-GB:W2F-003 #107:1>*
- (b) *OOh er missus - **what** a tizzy I'm **in** ! ! ! <ICE-IRE:W1B-004\$B>*
- (c) ***What** a lot of rush you've been **in** these last few days, eeh <ICE-JA:W1B-010#125:9>*
- (d) *Dev, **what** a sweet-man you turn **into** <ICE-JA:W2F-005#6:1>*
- (e) *good good lord **what** we were in **for** <ICE-NZ:S1B-042#46:1:J>*

Cleft RCs appeared with stranded (5.4a), pied-piped (5.4b), doubled (5.4c) and mismatching (5.4d) prepositions.

- (5.4) (a) *It is the resurrection **which** I am concerned **with** <ICE-IRE:W2A-007\$A>*
- (b) *The University of the West Indies offers the highest level of training within the Caribbean and it is **to this institution which** most persons aspire <ICE-JA:W1A-001#53:3>*
- (c) *And uh it's **from this guy Norhalis Shafik where** the first goal came **from** <ICE-SIN:S2A-009#131:1:A>*
- (d) *It is **in this ambitious list of commitments, drawn up after world-wide consultations and endorsed by over 150 governments, which Unicef is now calling on the world to rally behind** <ICE-EA-K:W2B026K>*

While these low token numbers did not allow for a multivariate statistical analysis, they support the claim made in section 3.1.1.2 that exclamative and information structure constructions are rather infrequent linguistic phenomena (compared to e.g. interrogative or finite RC constructions).

Table 5.2: Distribution of idiosyncratic tokens across VARIETY (absolute and relative frequencies)

VARIETY (VARIETY_TYPE)	Abs. freq.	Rel. freq.
CanE (V)	67	10.74%
BrE (V)	73	11.70%
IrE (V)	84	13.46%
NZE (V)	83	13.30%
JamE (IV)	81	12.98%
SgE (IV)	53	8.49%
KenE (III)	14	2.24%
TanE (III)	6	0.96%
HKE (III)	37	5.93%
IndE (III)	34	5.45%
NigE (III)	48	7.69%
PhilE (III)	44	7.05%
Total	624	100.00%

Another group that was excluded from the logistic regression analysis are the 624 idiosyncratic tokens that occur in the dataset (cf. table 5.2), i.e. *WH...like* and *WH...for* constructions. Although it is generally assumed that these strings only appear with stranded prepositions (cf. 3.1.2), the ICE data also contained three *WH...for* tokens with pied-piped prepositions (5.5).

- (5.5) (a) ***For what*** *she want to take?* <ICE-IND:S1B-079#189:1:B>
 (b) ***For what*** *was it created?* <ICE-IND:S2B-031#123:1:A>
 (c) *Uh* ***for what*** *you want to get* <ICE-SIN:S1A-086#271:2:E>

Even though this suggests that the constraint that requires these *WH...for/like* elements to be stranded may be looser in Englishes around the world than in Standard varieties, these tokens were excluded from the statistical analysis because they only occur in free RCs and interrogatives but not in finite RCs. Moreover, in order test the interaction effect of PP type and VARIETY/VARIETY_TYPE (cf. 5.2.2.1, 5.2.3.1), these tokens would have had to be grouped with other PP types to avoid level combinations for which the data contain no observations. This might have led to skewed results since the stranding preference of *WH...for/like* elements can be expected to be stronger than that of any other PP type (at least in Standard varieties). Although these *WH...for/like* tokens could thus

not be analyzed using mixed-effects models, they will be investigated with the help of HCFAs in section 5.3.

Finally, all 468 non-standard instances of preposition placement (cf. table 5.3), i.e. missing, doubling, mismatching, resumptive, extra and other tokens, were excluded from the logistic regression analysis. These tokens will be explored in section 5.5.

Table 5.3: Distribution of non-standard tokens across VARIETY (absolute and relative frequencies)

VARIETY (VARIETY_TYPE)	Abs. freq.	Rel. freq.
CanE (V)	9	1.92%
BrE (V)	7	1.50%
IrE (V)	20	4.27%
NZE (V)	11	2.35%
JamE (IV)	29	6.20%
SgE (IV)	68	14.53%
KenE (III)	37	7.91%
TanE (III)	54	11.54%
HKE (III)	71	15.17%
IndE (III)	74	15.81%
NigE (III)	56	11.97%
Phile (III)	32	6.84%
Total	468	100.00%

5.2.2 Preposition placement across Dynamic Model stages

5.2.2.1 Model selection

In order to test if there is a correlation between Dynamic Model stages and the strength of processing constraints and formality effects, a binomial mixed-effects model was fitted which included as many interactions of processing-related factors and stylistic variables with VARIETY_TYPE as possible. However, due to multicollinearity and data sparsity issues, not all of the variables that were coded (cf. 4.3) could be included in the maximal model.

As outlined in 3.1.2 and 4.4.1, associations between individual fillers and specific types of preposition placement can usually be explained by a correlation between filler type and clause type. As shown in figure 5.1, this effect also applies to the ICE data. While *which* is strongly associated with finite RCs, *what* is the prototypical filler in free RCs, embedded interrogatives and main clause interrogatives. Moreover, the use of *wh-*

ever elements is restricted to free RCs. To avoid multicollinearity issues, FILLER_TYPE was thus excluded from the model.

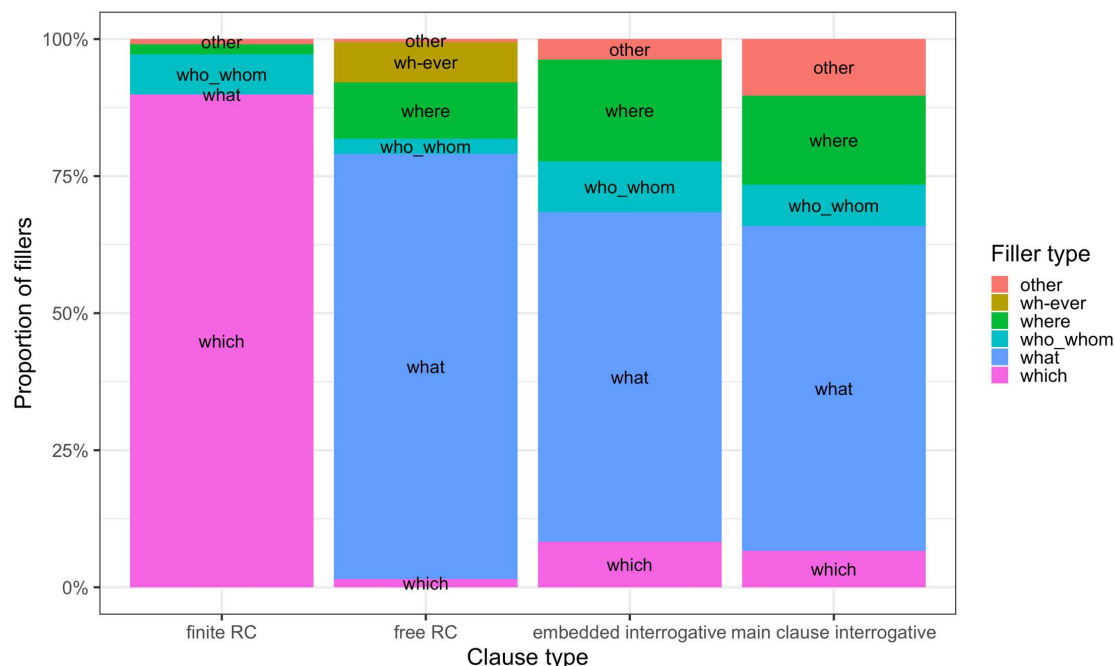


Figure 5.1: Distribution of FILLER_TYPE across CLAUSE_TYPE

Nevertheless, it is worth noting that there are 5 pied-piped *who*-tokens although P + *who* structures are normally considered ungrammatical in Standard varieties (cf. 3.1.2). These tokens can be found in all variety phases (but not in all varieties; cf. table 5.4), and they occur in both finite RCs (5.6a-d) and embedded interrogatives (5.6e).

- (5.6) (a) *which precluded making a lot of new friends (or any) in a totally new town, inheriting several of Steve's friends, most of **who** I don't connect with all that well* <ICE-CAN:W1B-012#48:2>
- (b) *because quite often <,> I believe <,> that politicians appoint people who **in who** they have no confidence* <ICE-JA:S1B-022#10:1:B>
- (c) *Marlu, I'll e-mail you the names and addresses of individuals and institutions **to who** we're addressing our letters of funding solicitation* <ICE-PHI:W1B-017#96:3>
- (d) *Philosophers **amongst who** we get the philosopher king okay* <ICE-EA-K:S2BLEC1K>
- (e) *I felt like shouting out "Fuck off" as I know **to who** it was directed but decided it would be more prudent to keep quiet* <ICE-GB:W1B-010 #56:2>

Table 5.4: Stranded and pied-piped *who*-tokens across varieties (absolute and relative frequencies)

VARIETY (VARIETY_TYPE)	<i>who</i>_stranded		<i>who</i>_pied-piped		<i>who</i>_total	
CanE (V)	18	94.74%	1	5.26%	19	100.00%
BrE (V)	28	96.55%	1	3.45%	29	100.00%
IrE (V)	31	100.00%	0	0.00%	31	100.00%
NZE (V)	44	100.00%	0	0.00%	44	100.00%
JamE (IV)	41	97.62%	1	2.38%	42	100.00%
SgE (IV)	14	100.00%	0	0.00%	14	100.00%
KenE (III)	5	83.33%	1	16.67%	6	100.00%
TanE (III)	2	100.00%	0	0.00%	2	100.00%
HKE (III)	9	100.00%	0	0.00%	9	100.00%
IndE (III)	2	100.00%	0	0.00%	2	100.00%
NigE (III)	15	100.00%	0	0.00%	15	100.00%
PhlE (III)	31	96.88%	1	3.12%	32	100.00%
Column total	240	97.96%	5	2.04%	245	100.00%

Table 5.5: Stranded and pied-piped *whom*-tokens across varieties (absolute and relative frequencies)

VARIETY (VARIETY_TYPE)	<i>whom</i>_stranded		<i>whom</i>_pied-piped		<i>whom</i>_total	
CanE (V)	4	20.00%	16	80.00%	20	100.00%
BrE (V)	1	2.38%	41	97.62%	42	100.00%
IrE (V)	3	7.50%	37	92.50%	40	100.00%
NZE (V)	2	11.11%	16	88.89%	18	100.00%
JamE (IV)	6	12.00%	44	88.00%	50	100.00%
SgE (IV)	8	22.22%	28	77.78%	36	100.00%
KenE (III)	4	13.79%	25	86.21%	29	100.00%
TanE (III)	1	3.85%	25	96.15%	26	100.00%
HKE (III)	5	33.33%	10	66.67%	15	100.00%
IndE (III)	4	13.79%	25	86.21%	29	100.00%
NigE (III)	6	23.08%	20	76.92%	26	100.00%
PhlE (III)	7	17.50%	33	82.50%	40	100.00%
Column total	51	13.75%	320	86.25%	371	100.00%

In addition, all varieties have stranded *whom*-tokens. Although the low frequencies make it impossible to draw definite conclusions, the token numbers given in tables 5.4 and 5.5 might be taken as an indication that stranding with *whom* is generally

more common than pied-piping with *who*, which is in line with the assumption that stranding is possible with both *who* and *whom* (cf. 3.1.2; Quirk et al. 1985: 370; Payne & Huddleston 2002: 465).

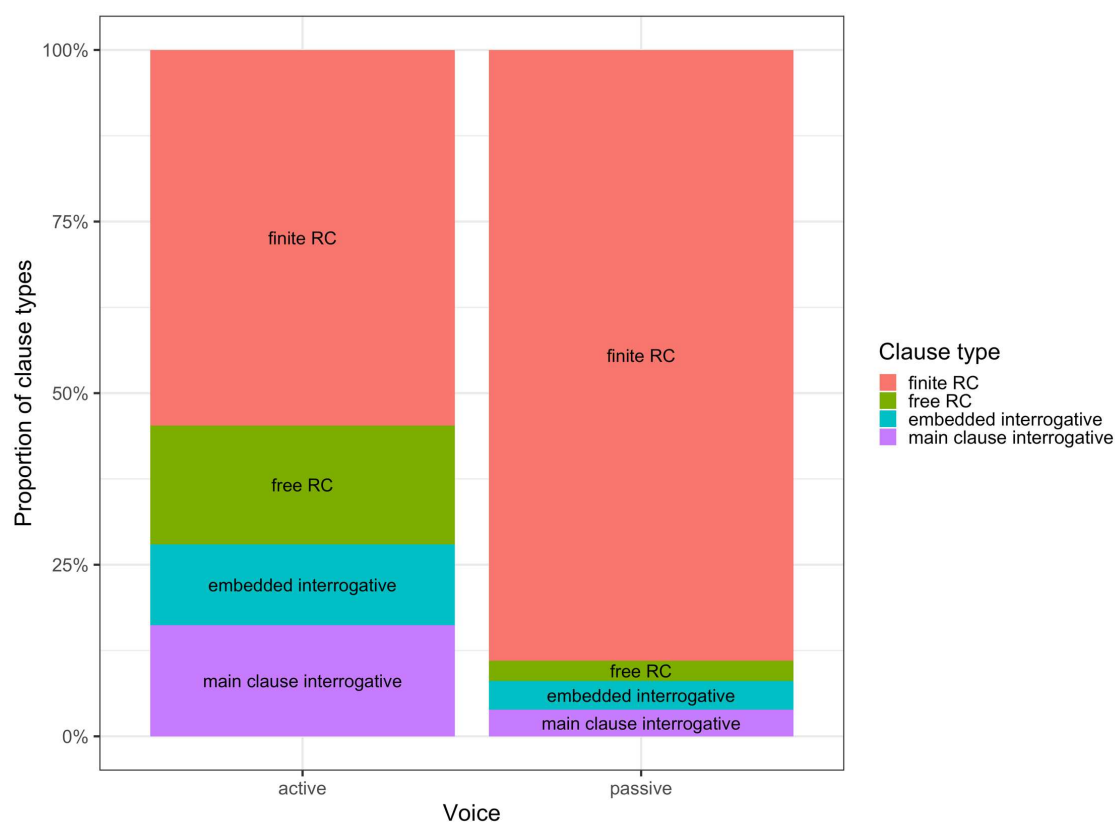


Figure 5.2: Distribution of *CLAUSE_TYPE* across *VOICE*

Returning to the fixed effects structure of the maximal model, another variable which had to be excluded to avoid high collinearity is *VOICE*. As shown in figure 5.2, the use of the passive is strongly associated with finite RCs. Hence, *VOICE* could have only been included as part of the three-way interaction *VOICE* * *CLAUSE_TYPE* * *VARIETY_TYPE*. However, as this complex interaction would have introduced many low frequency cells (cf. table A.2 in the Appendix), *VOICE* was excluded to avoid overfitting.

A third factor which could not be included in the maximal model is *COMPLEXITY_BRIDGE* because the data contained only few observations for the *complex_zero* (98, 1.02%) and *complex_that* (17, 0.18%) levels. The fact that the vast majority of the data have a syntactically simple bridging structure (9,496, 98.8%) suggests that speakers of all variety stages tend to avoid complex long-distance dependencies, which are extremely hard to process.

Table 5.6: Distribution of stranding and pied-piping in simple, complex_zero and complex_that tokens across VARIETY_TYPE (absolute and relative frequencies)

VARIETY_TYPE	COMPLEXITY_BRIDGE	PREPOSITION_PLACEMENT	Abs. Freq.	Rel. Freq.
III	complex_that	stranding	0	0.00%
		pied-piping	4	100.00%
	complex_zero	stranding	20	55.56%
		pied-piping	16	44.44%
	simple	stranding	1,494	37.57%
		pied-piping	2,483	62.43%
IV	complex_that	stranding	3	75.00%
		pied-piping	1	25.00%
	complex_zero	stranding	14	66.67%
		pied-piping	7	33.33%
	simple	stranding	923	51.62%
		pied-piping	865	48.38%
V	complex_that	stranding	1	11.11%
		pied-piping	8	88.89%
	complex_zero	stranding	28	68.29%
		pied-piping	13	31.71%
	simple	stranding	1,889	50.63%
		pied-piping	1,842	49.37%

Table 5.7: HCFA results for VARIETY_TYPE/VARIETY * VERB_BRIDGE (non-significant associations are not shown)

VARIETY_TYPE	VARIETY	VERB_BRIDGE	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
.	.	think	50	5.6522	347.9579	>	1.72131102043953E-32	***	0.357
.	.	say	29	5.6522	96.4438	>	7.30936344635289E-12	***	0.188

Although the low token numbers made it impossible to draw conclusions about the preposition placement preferences of constructions with syntactically complex bridging structures (cf. table 5.6²⁴), it was decided to take a closer look at the

²⁴ The frequencies in table 5.6 include all tokens except non-standard types of preposition placement.

complex_zero and complex_that tokens in the dataset. In section 3.2.4, it was pointed out that semantically weak and unspecific bridging verbs are easier to process than semantically more complex and more specific words (Kluender 1990: 199; Hawkins 2004: 201). In order to examine the bridging verbs in the present data in more detail, an HCFA of all complex_zero and complex_that tokens was conducted to test whether certain bridging verbs are preferred in certain varieties or Dynamic Model stages. No bridging verb turned out to be significantly associated with a specific variety or variety phase, which can be explained by the low token numbers of the individual levels of VERB_BRIDGE. However, the HCFA revealed that, across all varieties, the verbs *think* (5.7) and *say* (5.8) occur with a higher frequency than expected by chance (cf. table 5.7).

- (5.7) (a) *lets see what he **thinks** this match is going to be like* <ICE-IND:S2A-004#11:1:A>
- (b) *Anyway we had a discussion on the "misterio escondido" of "El celoso extremeno" which I still have not found and **think** that we should not waste our time looking for* <ICE-JA:W1B-005#57:4>
- (5.8) (a) *No doubt it will what it will create some pressure on her going through this ordeal of giving evidence in court against people whom she **say** she's very frightened of* <ICE-HK:S2A-066#105:1:A>
- (b) *the bridge we now we want to cross now is the modalities for the election in which we are **saying** that we want international observers to come and monitor the registration of voters* <ICE-NIG:bint_05:a450>

In accordance with the assumptions of Kluender (1990: 199) and Hawkins (2004: 201), the verbs *think* and *say* account for 68.7% (79/115) of all complex_zero and complex_that tokens in the data. As a result of the high processing cost involved in long-distance dependencies, speakers of all varieties thus seem to use constructions with complex bridging structures primarily with a small set of unspecific verbs that reduce semantic processing.

While the variables FILLER_TYPE, VOICE and COMPLEXITY_BRIDGE were excluded from the model, other factors had to be recoded to avoid combinations of factor levels with very low frequencies. With respect to the linguistic factors, several levels of the fine-grained PP_TYPE classification described in 4.3.1 had to be conflated to avoid data sparsity issues (which were particularly problematic with respect to the interaction VARIETY_TYPE * PP_TYPE; cf. table 5.8) and to reduce collinearity (as some of the PP types have very similar preposition placement preferences).

Table 5.8: Distribution of PP_TYPE across VARIETY_TYPE (absolute and relative frequencies)

PP_TYPE	Phase III		Phase IV		Phase V		Total	
V-X-P	64	1.67%	40	2.39%	120	3.46%	224	2.50%
prepositional X	591	15.43%	323	19.27%	546	15.75%	1,460	16.27%
obligatory complement	221	5.77%	142	8.47%	294	8.48%	657	7.32%
optional complement	934	24.38%	419	25.00%	847	24.43%	2,200	24.52%
movement	238	6.21%	122	7.28%	214	6.17%	574	6.40%
accompaniment	48	1.25%	25	1.49%	41	1.18%	114	1.27%
subcategorized PP	46	1.20%	24	1.43%	38	1.10%	108	0.82%
affected location	35	0.91%	13	0.78%	26	0.75%	74	0.82%
direction	10	0.26%	0	0.00%	16	0.46%	26	0.29%
location	172	4.49%	83	4.95%	176	5.08%	431	4.80%
time	129	3.37%	57	3.40%	145	4.18%	331	3.69%
duration/frequency	102	2.66%	37	2.21%	99	2.86%	238	2.65%
means/instrument	212	5.53%	56	3.34%	124	3.58%	392	4.37%
agent	5	0.13%	0	1.97%	1	0.03%	6	0.07%
cause/reason/purpose/result	110	2.87%	33	0.00%	43	1.24%	186	2.07%
respect	596	15.56%	178	10.62%	416	12.00%	1,190	13.26%
manner	217	5.66%	85	5.07%	232	6.69%	534	5.95%
degree	101	2.64%	39	2.33%	89	2.57%	229	2.55%
Column total	3,831	100%	1,676	100%	3,467	100%	8,974	100%

Table 5.9: Recoding of PP_TYPE

Factor level	PP type
lexicalized	VXP
	prepositional X
	obligatory complement with <i>be</i>
complement_like	obligatory complement without <i>be</i>
	optional complement
	movement
	accompaniment
adjunct_like	subcategorized PP
	affected location
	direction
	location
	time
	duration/frequency
	means/instrument
	agent
	cause/reason/purpose/result
respect_manner_degree	respect
	manner
	degree

As shown in table 5.9, PP types with the strongest association between verb and preposition, i.e. prepositional verbs and VXP elements, were combined into the “lexicalized” level. Obligatory complements with *be*, which have been found to strongly prefer stranding in BrE, were also included in this group. The “complement_like” group contains optional complements, obligatory complements without *be*, as well as accompaniment and movement PPs, whose stranding preference has been shown to be comparable to that of optional complements (cf. 3.2.2). The remaining adjunct PPs were divided into two levels: Due to the almost categorical pied-piping preference of respect, manner and degree PPs, these PP types were combined into one “respect_manner_degree” level. All other adjunct PPs as well as subcategorized PPs were assigned to the level “adjunct_like”.

Interestingly, the present data contained 8 stranded respect PPs although Hoffmann’s (2011: 160) corpus study found a categorical pied-piping preference for respect, manner and degree PPs in ICE-GB and ICE-EA-K. These stranded respect PP

tokens come from all variety phases and they include free RCs (5.9a), embedded interrogatives (5.9b) and finite RCs (5.9c).

- (5.9) (a) *I think everybody is a good person you know depends on **what perspective you're looking at the person from*** <ICE-NIG:con_32:a205>
- (b) *I don't know **what they're scoring on*** <ICE-IRE:S2A-003\$C>
- (c) *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>

Note that (5.9a) was coded as a respect PP because here *look at* is used in a metaphorical sense with the meaning ‘examine’ or ‘think about’, and the PP (*from a perspective*) establishes a point of reference against which the examination of the person is evaluated. Example (5.9b) is supposed to mean ‘I don’t know on what basis the judges decide how many points a boxer gets in a fight’, where the PP establishes a reference point for the decisions of the judges. In (5.9a) and (5.9b), the use of stranded prepositions might be due to fact that the construction types per se, i.e. free RCs and interrogatives, strongly favor stranding. However, in (5.9c), the stranded respect PP occurs in a finite RC, i.e. in a construction type which, due to its complexity, normally strongly favors pied-piping. As this suggests that stranding with respect, manner and degree adjunct PPs is not entirely impossible in World Englishes, the respect, manner and degree PP tokens were not excluded from the logistic regression analysis.

Another variable which had to be recoded was `CLAUSE_TYPE`. Recall that Hoffmann (2011: 156) found that only finite RCs are affected by formality (cf. 3.1.1.1). In order to test this effect, the three-way interaction `CLAUSE_TYPE * FORMALITY * VARIETY_TYPE` had to be included in the logistic regression model. However, as the token numbers of pied-piped free RCs and pied-piped interrogatives in informal contexts are very low, `CLAUSE_TYPE` had to be reduced to a binary distinction between the levels “freeRC_interrogative” and “finiteRC”, which also helped to reduce collinearity as both free RCs and interrogatives have a very strong preference for preposition stranding. Importantly, however, table 5.10 shows that the present data contain pied-piped free RCs, which supports Hoffmann’s (2011: 262) claim that “pied-piping should not be completely ruled out in free relative clauses”. These pied-piped free RC tokens will be explored in section 5.3.4.

Table 5.10: Distribution of stranding and pied-piping per *CLAUSE_TYPE* across *VARIETY_TYPE* (absolute and relative frequencies)

VARIETY_ TYPE	CLAUSE_TYPE	PREPOSITION_ PLACEMENT	Abs. freq.	Rel. freq.
III	embedded	stranding	271	72.46%
	interrogative	pied-piping	103	27.54%
	finite RC	stranding	263	10.63%
		pied-piping	2,211	89.37%
	free RC	stranding	460	92.56%
		pied-piping	37	7.44%
	main clause	stranding	345	70.99%
	interrogative	pied-piping	141	29.01%
IV	embedded	stranding	179	91.33%
	interrogative	pied-piping	17	8.67%
	finite RC	stranding	154	16.04%
		pied-piping	806	83.96%
	free RC	stranding	252	96.92%
		pied-piping	8	3.08%
	main clause	stranding	220	84.62%
	interrogative	pied-piping	40	15.38%
V	embedded	stranding	327	89.10%
	interrogative	pied-piping	40	10.90%
	finite RC	stranding	243	12.06%
		pied-piping	1,772	87.94%
	free RC	stranding	562	98.60%
		pied-piping	8	1.40%
	main clause	stranding	472	91.65%
	interrogative	pied-piping	43	8.35%

Table 5.11: Distribution of *PHRASE_TYPE* across *VARIETY_TYPE* (absolute and relative frequencies)

VARIETY_TYPE	AdjP		VP		NP		Total	
III	148	3.86%	3,496	91.26%	187	4.88%	3,831	100%
IV	74	4.42%	1,540	91.89%	62	3.70%	1,676	100%
V	114	3.29%	3,221	92.90%	132	3.81%	3,467	100%
Column total	336	3.74%	8,257	92.01%	381	92.01%	8,974	100%

The only categorical linguistic variable that has not been mentioned yet is *PHRASE_TYPE*, whose distribution across variety phases is shown in table 5.11.

Although the token numbers of AdjP- and NP-contained PPs are considerably lower than those of VP-embedded PPs, the frequencies were sufficient for the regression analysis.

Coming to the extralinguistic variables, CATEGORY and TOPIC had to be recoded as they turned out to be correlated. In order to explore the distribution of topics across ICE text types, the 20 topics obtained from the topic model were categorized as “informal”, “medium” and “formal” (Mußemann 2024: 13). Based on Koch & Oesterreicher’s (2012: 450) model of “language of immediacy” and “language of distance”, which conceptualizes the relationship between spoken and written language by taking into account both conditions of communication and strategies of verbalization, the personal topic, which should be characterized by a very involved, affective style and a low degree of planning, is considered the most informal topic, i.e. the topic marked by the highest degree of immediacy. Formal topics, such as administration, science or philosophy, are defined by features of the language of distance, such as detachment, objectivity and information density. The medium level includes topics which are informational but which also affect speakers personally to a certain extent. For example, with regard to education or society, it turned out that speakers do not just present abstract information but that they also relate personal experiences in terms of e.g. schooling or family life. Likewise, sports texts typically do not merely provide information in an objective way but they also include emotional comments and personal opinions. Hence, these topics should be marked by a greater degree of involvement and expressivity than the most abstract and informational topics.²⁵ An overview of all topics categorized according to formality is given in table 5.12.

Table 5.12: Topics categorized according to formality (Mußemann 2024: 29)

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

²⁵ This classification is admittedly somewhat *ad hoc* and may need revision in future studies.

At the same time, the ICE text types were also categorized into the levels “informal”, “medium” and “formal” (cf. table 5.13). Considering Hoffmann’s (2011: 83) claim that “a simple spoken/written dichotomy may not suffice” to investigate the effect of stylistic variables, this three-level distinction takes into account that both spoken and written language can be formal or informal. Private dialogues are considered informal as they have been shown to be the ICE category which is characterized by the most involved, interactive and casual style (Xiao 2009: 436; Bohmann 2019: 107). Private correspondence texts are also assigned to the informal level since these texts should be marked by a greater degree of personal involvement than all other written text types (Bohmann 2019: 108). In contrast, the business correspondence section and printed texts, i.e. academic writing, non-academic writing, reportage, instructional writing and persuasive writing, are classified as formal since these categories should be very informative and elaborate in style (Xiao 2009: 436). In accordance with Xiao’s (2009: 436) finding that novels and stories “are more akin to spoken registers (possibly because of the heavy use of fictional dialogues)”, the only printed section not included in the formal level is creative writing, which is categorized as medium. Student essays and examinations, which make up the category non-printed non-professional writing, are also assigned to the medium level since these texts have been found to be more casual and less elaborative than the printed, edited texts (Xiao 2009: 436–437; Bohmann 2019: 107). With respect to the spoken categories, scripted monologues, i.e. e.g. speeches or broadcast talks, are classified as formal because, according to Bohmann (2019: 108), “the scripted monologue genre is not significantly different from most written ones”, which can be explained by the fact that this genre “is only medially oral, but conceptionally decidedly literate”. Public dialogues and unscripted monologues belong to the medium level because, in terms of formality, these texts occupy an intermediate position between the very involved, informal private dialogues and the very informational, formal scripted monologues (Xiao 2009: 436–437; Bohmann 2019: 107–108). The category broadcast news is also classified as medium since these texts can be considered a mixture of dialogue and monologue (Hoffmann 2011: 114). Finally, with respect to the categories which only exist in ICE-EA, the written as spoken texts are coded as medium because they include relatively spontaneous spoken language from parliament or legal cross-examinations. At the same time, the legal presentations from ICE-EA, i.e. “formal written manuscripts which are to be read out” (Hoffmann 2011: 150), are assigned to the formal level.

Table 5.13: ICE text types categorized according to formality

	Informal	Medium	Formal
Spoken	Private dialogue	Public dialogue Unscripted monologue Broadcast news	Scripted monologue
Written	Private correspondence	Non-professional writing Creative writing	Academic writing Non-academic writing Instructional writing Persuasive writing Reportage Business correspondence
ICE-EA only		Written as spoken	Legal presentations

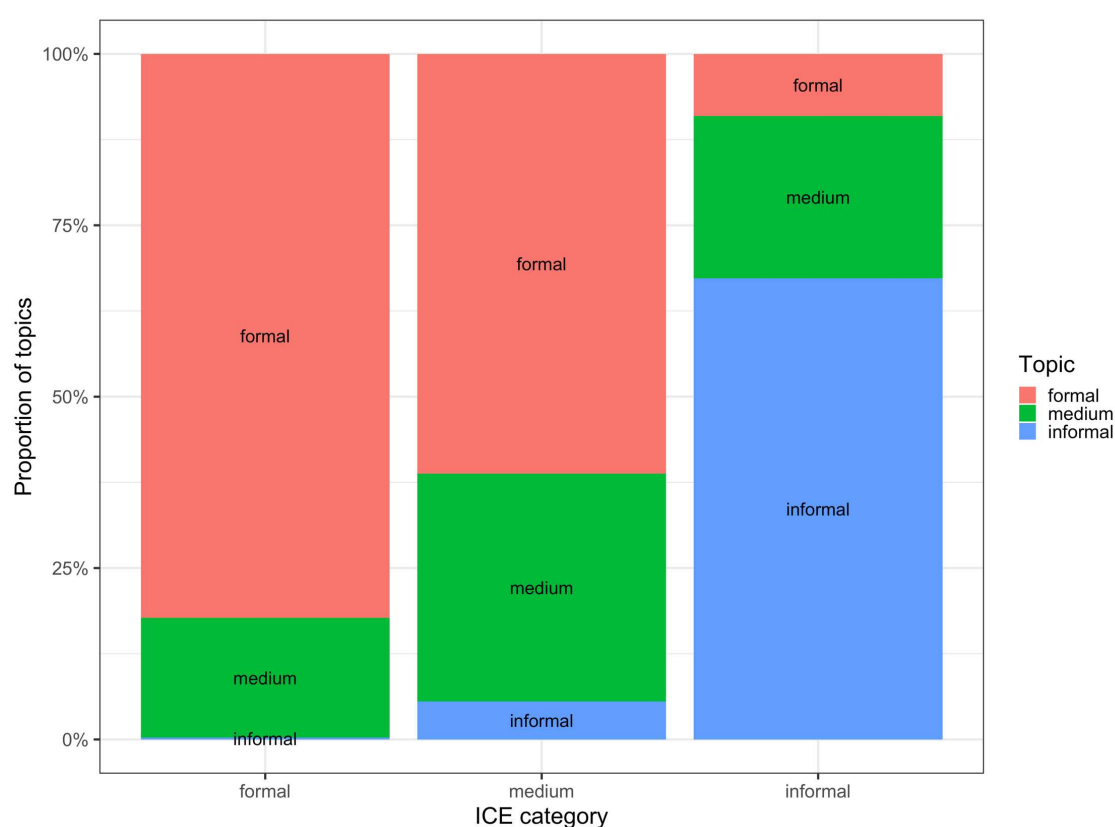


Figure 5.3: Distribution of topics across ICE categories

As shown in figure 5.3, the topics and the ICE text types are largely correlated. While medium and formal topics, such as science, economy or national development, tend to occur in medium or formal text types, the informal topic clearly dominates the informal categories. However, there also some noteworthy exceptions.

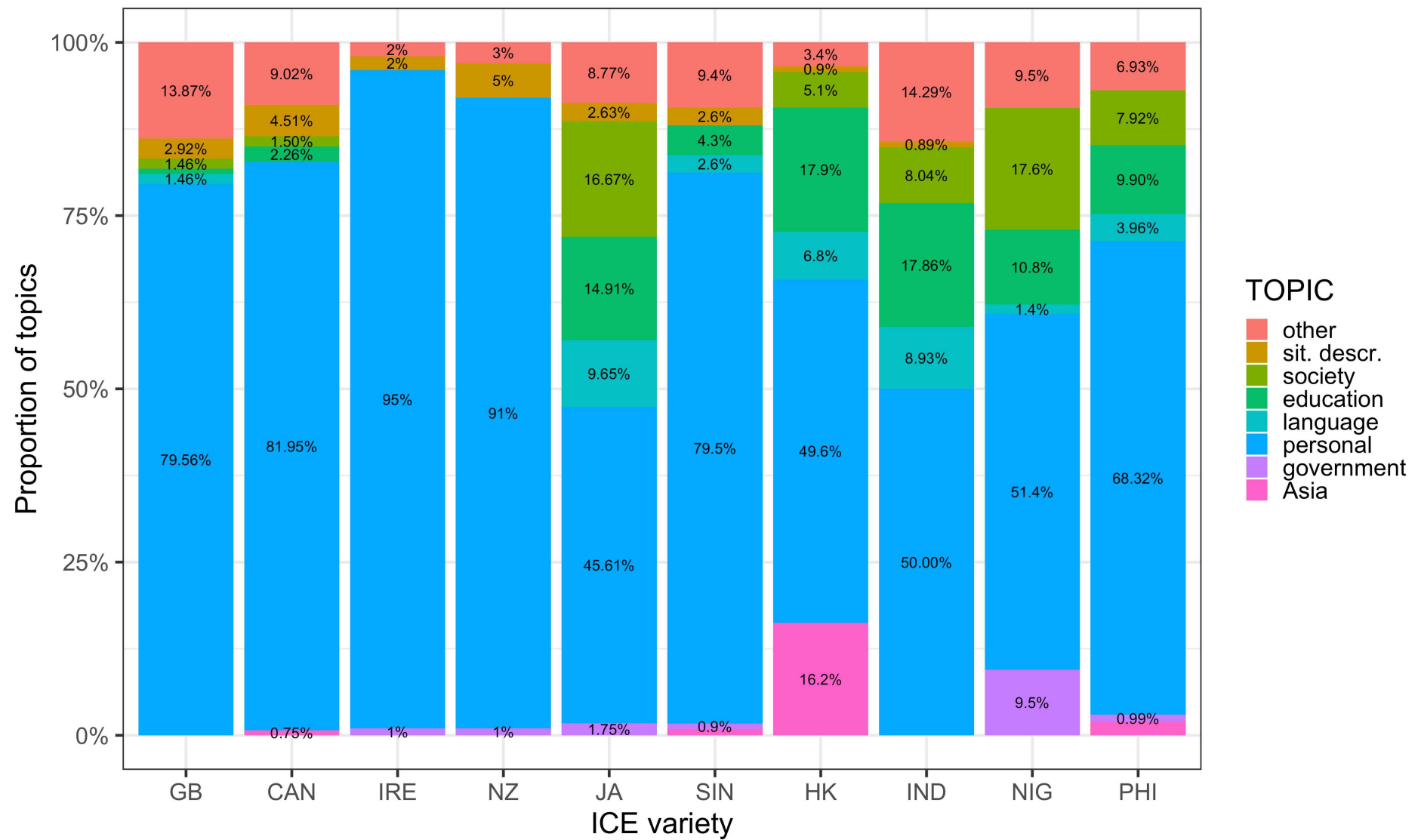


Figure 5.4: Topics per ICE variety in private dialogues. ICE-EA is excluded because of its different corpus design.

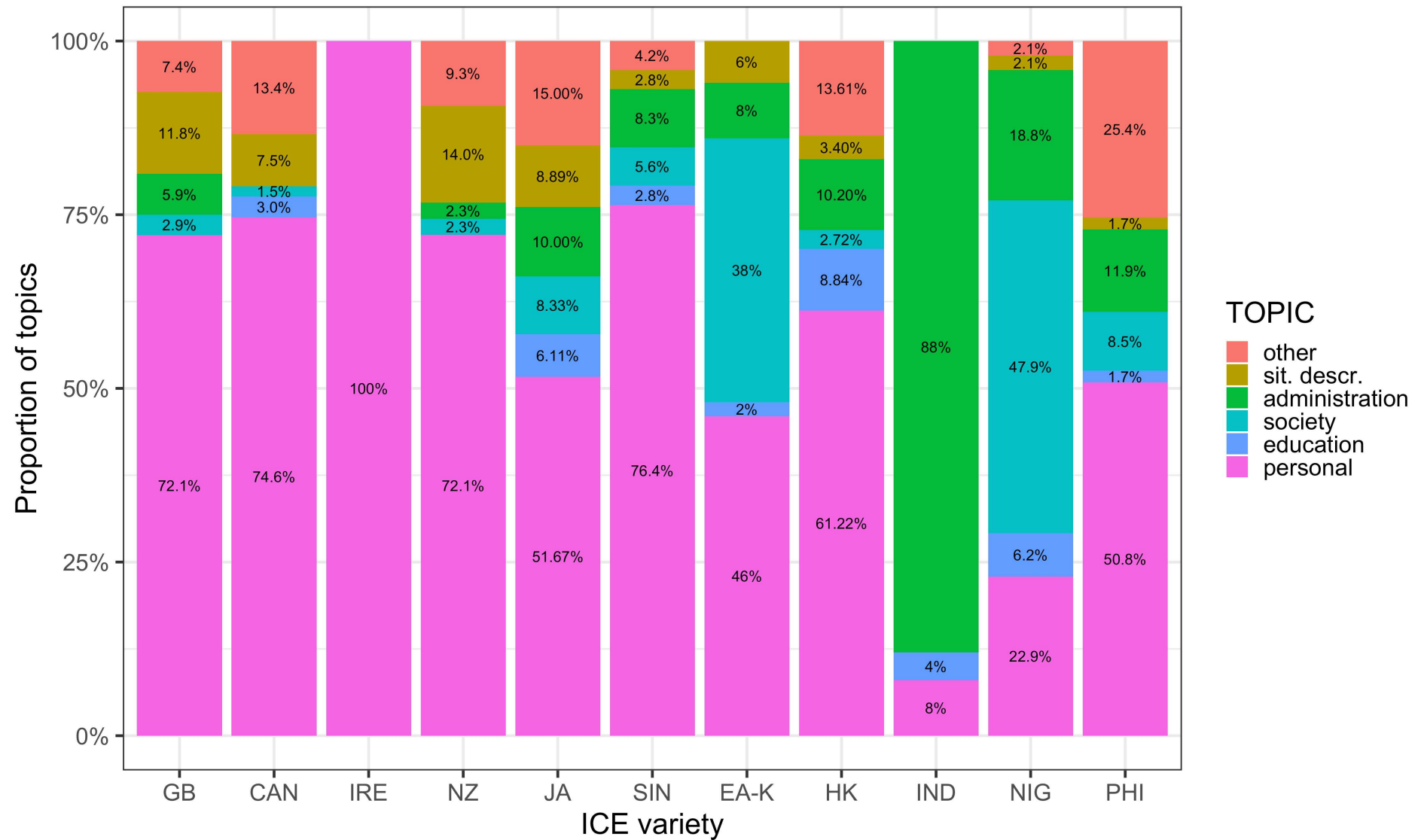


Figure 5.5: Topics per ICE variety in private correspondence. ICE-Tanzania is excluded because it does not contain social letters (adapted from Mußemann 2024: 30).

First of all, the private dialogue sections of less advanced varieties seem to be characterized by more formal topics than those of varieties at higher Dynamic Model stages. The personal topic, which is considered the most informal topic, makes up more than 75% of all private dialogues in ICE-GB, ICE-CAN, ICE-IRE, ICE-NZ and ICE-SIN, whereas the S1A-files of ICE-JA, ICE-HK, ICE-IND, ICE-NIG and ICE-PHI have higher proportions of medium and formal topics, such as education, society or government (cf. figure 5.4). Since PCEs are typically used in formal domains (cf. 2.3), it might have felt unnatural for these corpus speakers to use English for personal matters during the ICE recordings. At the same time, SgE is increasingly used as a home language (Wong 2014: 4), which may explain why ICE-SIN patterns with the Stage-V varieties. The assumption that ICE-HK may have a more formal private dialogue section than the other ICE components (cf. 4.1) thus seems to apply to all PCEs included in this study (with the exception of SgE).

Moreover, a closer look at the informal texts dealing with formal topics reveals that ICE-IND strongly differs from all other varieties with regard to the private correspondence texts. While private correspondence files usually tend to be dominated by personal topics, in ICE-IND these texts mainly focus on administrative issues (cf. figure 5.5). Consequently, the private correspondence files of ICE-IND can be considered more formal than those of the other ICE components (also see Mußemann 2024: 13).

To avoid high collinearity, it was decided to recode CATEGORY and TOPIC into a single variable FORMALITY, which consists of the two levels “more_formal” and “less_formal” (cf. table 5.14). Note that a binary distinction was implemented to ensure that the frequencies of all level combinations were sufficiently high for the testing of the complex three-way interaction `CLAUSE_TYPE * FORMALITY * VARIETY_TYPE`, which would not have been possible with a three-level distinction. As I argued elsewhere (Mußemann 2024: 13), this approach has many advantages over a formality classification based on the ICE categories only as it takes into account differences that exist between the various ICE varieties with respect to formality and topic despite the shared corpus design.

Consequently, private correspondence texts about formal topics were classified as more formal, whereas social letters dealing with other topics were assigned to the less formal level. The private dialogues were all categorized as less formal due to the casual and interactive style that is typical of these texts (Xiao 2009: 436; Bohmann 2019: 107). In addition, with respect to public dialogues and unscripted monologues, which had been

classified as medium in table 5.13, a nuanced formality classification was implemented that is based on topics: While texts dealing with informal (i.e. personal) topics were considered less formal, public dialogues and unscripted monologues about other topics (e.g. broadcast discussions about education or parliamentary debates about the economy) were categorized as more formal. In addition, some text types (cf. table 5.14) were assigned to the more formal level regardless of topics because they represent edited texts or because they have been shown to be more informational and elaborate in style than the texts of the less formal level (Xiao 2009: 436–437; Bohmann 2019: 107–108). The distribution of the variable FORMALITY across variety phases is shown in table 5.15.

Table 5.14: Levels of FORMALITY

Less_formal	More_formal
Private dialogue	Private correspondence about formal topics
Private correspondence about informal/medium topics	Private dialogues/unscripted monologues about medium/formal topics
Public dialogues/unscripted monologues about informal topics	Regardless of topic:
	- Scripted monologues
	- Broadcast news
	- Non-professional writing
	- Business correspondence
	- Academic writing
	- Non-academic writing
	- Reportage
	- Instructional writing
	- Persuasive writing
	- Creative writing
	- Only ICE-EA: Written as spoken, Legal presentations

Table 5.15: Distribution of FORMALITY across VARIETY_TYPE (absolute and relative frequencies)

VARIETY_TYPE	more_formal		less_formal		Total	
III	3,285	85.75%	546	14.25%	3,831	100%
IV	1,316	78.52%	360	21.48%	1,676	100%
V	2,662	76.78%	805	23.22%	3,467	100%
Column total	7,263	80.93%	1,711	19.07%	8,974	100%

With respect to the numeric predictors, the factors COMPLEXITY, FILLER_LENGTH, FRQUENCY_FILLER, and FREQUENCY_PREPOSITION were log-transformed, centered and standardized. Regarding FILLER_LENGTH, it should be noted that the data contain only 245 tokens with fillers consisting of three or more words (cf. table 5.16), while the vast majority of the data (8,014 tokens) have simple fillers consisting of one word. Despite this skewed distribution (cf. figure A.3 in the Appendix), which cannot be corrected by the log-transformation, it was not an option to convert FILLER_LENGTH into a categorical variable. For instance, a categorical variable with a three-level distinction “short”, “medium” and “long” would have had only very few observations for the levels medium and long, which would not have permitted the testing of any interaction effects. Moreover, for the interpretation of the results, it should be kept in mind that most complex filler phrases occur in interrogatives (cf. figure A.4 in the Appendix), and that they are particularly frequent in constructions with frequency and degree PPs (cf. figure A.5 in the Appendix).

Table 5.16: Tokens with FILLER_LENGTH >= 3 across variety phases (absolute and relative frequencies)

VARIETY_TYPE	Abs. freq.	Rel. freq.
III	94	38.4%
IV	69	28.2%
V	82	33.5%
Total	245	100%

After these preparatory steps, a maximal model was fitted which included four three-way interactions: CLAUSE_TYPE * FORMALITY * VARIETY_TYPE tests Hoffmann’s (2011: 156) finding that formality mostly affects finite RCs, while COMPLEXITY * CLAUSE_TYPE * VARIETY_TYPE accounts for the fact that finite RCs tend to be longer than free RCs and interrogatives (cf. figure A.6 in the Appendix). Furthermore, the three-way interaction FREQUENCY_FILLER * FORMALITY * VARIETY_TYPE was included because more frequent fillers tend to be slightly more frequent in informal texts than in formal ones (cf. figure A.7 in the Appendix). This might be due to the frequent occurrence of frequent fillers such as *what* in constructions that are typical of informal contexts such as *What is it all about?* or *What are you up to?*. Likewise, FREQUENCY_PREPOSITION * FORMALITY * VARIETY_TYPE was included to account for the fact that more formal texts contain slightly more frequent prepositions than informal ones (cf. figure A.8 in the Appendix). This effect could result

from the frequent use of *in* in the P + WH string *in which* in formal finite RCs. Apart from these three-way interactions, the maximal model tested the two-way interactions VARIETY_TYPE * PP_TYPE, VARIETY_TYPE * FILLER_LENGTH, and VARIETY_TYPE * PHRASE_TYPE. To investigate whether potential associations between specific clause types and specific types of preposition placement are due to the fact that certain PP types occur more frequently in some clause types than in others, it would have been helpful to also test the three-way interaction CLAUSE_TYPE * PP_TYPE * VARIETY_TYPE. However, this was not possible due to data sparseness. Likewise, it would have been useful to include the interaction FILLER_LENGTH * PP_TYPE * VARIETY_TYPE to account for the fact that long filler phrases tend to appear in constructions with adjunct PPs, which normally tend to be pied-piped. However, due to the infrequent occurrence of long filler phrases, this interaction would have led to large standard errors and overfitting.

Turning to the random effects, the maximal model included random intercepts for FILE_ID and PREPOSITION (cf. 4.4.1). Prepositions for which the data contain less than 5 tokens (N = 38) were pooled into one group “other”. Although there is a large number of FILE_ID levels with only one observation (N = 1,694), these levels were not pooled together as this would have created one level with tokens from 1,694 speakers. Even though the estimates of the random effects variance of FILE_ID may thus be inaccurate, the large number of FILE_ID levels for which only one observation exists should not have a negative impact on the fixed effects estimates (Meteyard & Davies 2020). A more complex random effects structure was not implemented to avoid convergence issues.

The final model structure, from which all non-significant random and fixed effects were removed using the backward elimination procedure outlined in 4.4.1, is shown in (5.10):

$$\begin{aligned}
 (5.10) \text{ PREPOSITION\_PLACEMENT} \sim & \\
 & \text{VARIETY\_TYPE} + \text{CLAUSE\_TYPE} + \text{FORMALITY} + \\
 & \text{COMPLEXITY} + \text{PP\_TYPE} + \text{PHRASE\_TYPE} + \\
 & \text{FILLER\_LENGTH} + \text{FREQUENCY\_FILLER} + \\
 & \text{CLAUSE\_TYPE:FORMALITY} + \\
 & \text{CLAUSE\_TYPE:COMPLEXITY} + \\
 & \text{VARIETY\_TYPE:CLAUSE\_TYPE} + \\
 & \text{VARIETY\_TYPE:FORMALITY} + \\
 & \text{VARIETY\_TYPE:PP\_TYPE} +
 \end{aligned}$$

$$(1 \mid \text{FILE\_ID}) + (1 \mid \text{PREP})$$

This model is highly significantly better ($p < 0.001$) than the corresponding null model without any fixed effects. It has a marginal R^2 of 0.8 and a conditional R^2 of 0.91. All adjusted GVIFs are clearly below 10, which indicates that multicollinearity should not be a problem. The condition number K is 20.48, which also suggests that there is no harmful collinearity (cf. 4.4.1). The classification accuracy of the model is 96.02% (precision = 95.8%, recall = 97.44%), which is highly significantly better than a model that always choses pied-piping ($p < 0.001$). The C -index is 0.99, which indicates that the model has outstanding predictive capacities.

5.2.2.2 Results

Table 5.17: Variance estimates and standard deviations of random effects (of the final model focusing on interactions involving *VARIETY_TYPE*)

Groups	Variance	SD
FILE_ID	3.214	1.793
PREPOSITION	1.003	1.001

The random intercept estimates for FILE_ID and PREPOSITION in table 5.17 indicate that both individual speakers and individual prepositions have idiosyncratic preferences for preposition stranding and pied-piping. Figure 5.6, in which positive estimates indicate a stranding preference and negative estimates illustrate a pied-piping preference, visualizes the intercept adjustments for the various prepositions. *Since*, *around* and *among* emerge as the prepositions with the strongest pied-piping probability, while *against*, *through* and *about* strongly favor stranding. The pied-piping preference of *since*, which is in line with previous research (e.g. Quirk et al. 1985: 817), might be due to the fact that *since* typically appears in frequency PPs, which strongly favor pied-piping. *Among* and *around* may favor pied-piping because they frequently form part of respect PPs ((5.11) and (5.12)), which also tend to become uninterpretable with stranded prepositions (cf. 3.2.2).

(5.11) *then produce a documented picture of the most significant environmental developments **around which** the organization must formulate its future goals, strategies, structures and systems* <ICE-HK:W2A-032#47:1>

(5.12) *Gildas is of the view that the ruin of Britain began with the islands throwing up of usurpers, **among whose rebellions** that of magnus maximus*

Magnus Maximus was the most significant and dangerous to Britains security <ICE-GB:W1A-003 #110:2>

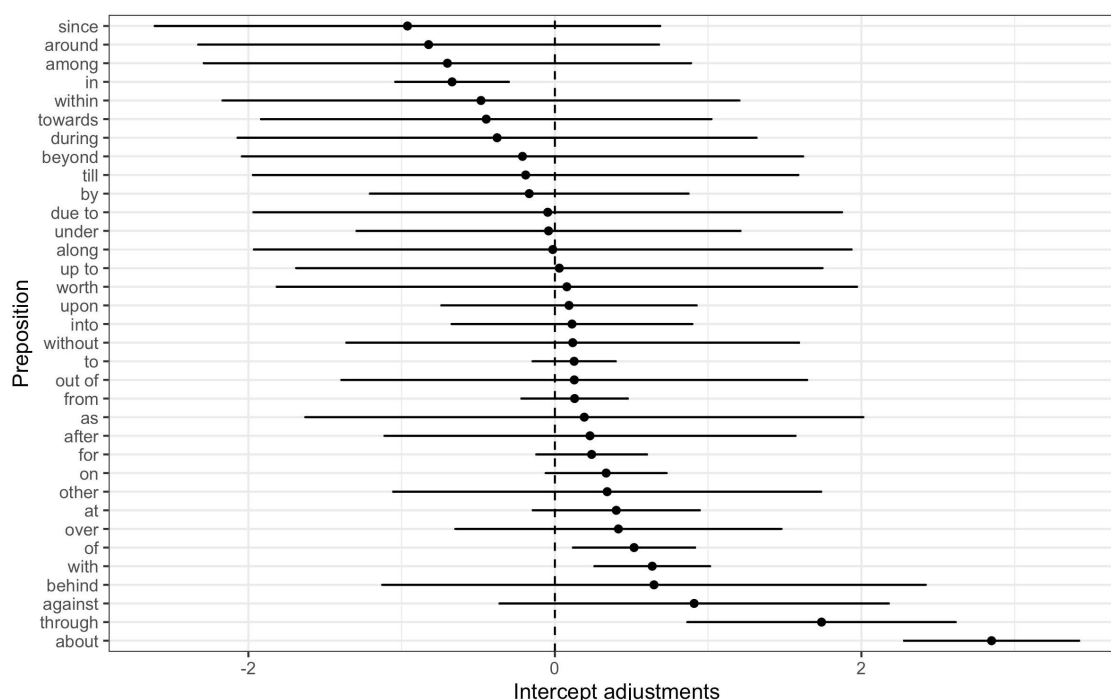


Figure 5.6: Intercept adjustments of PREPOSITION and Wald confidence intervals (of the final model focusing on interactions involving VARIETY_TYPE)

The strong stranding preference of *about* could be due to its frequent occurrence in *what-SUBJ-BE-about*-constructions (5.13), and the fact that verbs such as *speak* (5.14a), *write* (5.14b) or *talk* (5.14c) are typically followed by optional complement PPs that involve *about* (cf. 3.2.2).

(5.13) *That's not **what** linguistics is **about*** <ICE-GB:S1A-038 #313:1:C>

(5.14) (a) *you don't know **what** you're speaking **about*** <ICE-GB:S1A-038 #139:1:A>

(b) *Emma **what** are you going to write **about*** <ICE-GB:S1B-007 #207:1:A>

(c) ***What** are we talking **about*** <ICE-GB:S1A-030 #190:1:B>

The idiosyncratic stranding effect of *through* can be explained by *go-through*- (5.15a) and *be-through*-tokens (5.15b) in which *through* is used in a metaphorical sense to express the meaning ‘suffer’ or ‘experience’.

(5.15) (a) *we are involved in the case and we know **what** uh people are going **through*** <ICE-CAN:S1B-048#109:1:B>

- (b) *i can only imagine **what** you've been **through** in the past many months* <ICE-CAN:W1B-008#3:1>

Whether some of these *WH...about* or *WH...through* constructions are particularly frequent in a specific variety or variety type and can thus be said to be entrenched in the construction networks of certain varieties will be examined in section 5.3.

Table 5.18: Coefficients of significant main effects (of the final model focusing on interactions involving *VARIETY_TYPE*). Model predictions are for preposition stranding.

Predictors	<i>b</i>	SE	<i>z</i>	<i>p</i>
(Intercept)	6.9625	0.5722	12.1671	<0.001
VARIETY_TYPE				
V > III	-3.4189	0.5125	-6.6713	<0.001
V > IV	-1.4747	0.6137	-2.403	<0.05
CLAUSE_TYPE				
freeRC_interrogative > finiteRC	-4.6162	0.4332	-10.6564	<0.001
FORMALITY				
less formal > more formal	-1.9719	0.3933	-5.0139	<0.001
COMPLEXITY (logged)				
one-unit increase	-1.0857	0.1806	-6.0104	<0.001
PP_TYPE				
lexicalized > complement_like	-0.8535	0.253	-3.3739	<0.001
lexicalized > adjunct_like	-3.539	0.3585	-9.8716	<0.001
lexicalized > respect_manner_degree	-7.5719	0.7616	-9.9422	<0.001
PHRASE_TYPE				
VP > AdjP	-0.1663	0.245	-0.679	0.4971
VP > NP	-1.3098	0.2882	-4.5448	<0.001
FILLER_LENGTH (logged)				
one-unit increase	-0.2532	0.069	-3.6726	<0.001
FREQUENCY_FILLER (logged)				
one-unit increase	0.3497	0.0546	6.4049	<0.001

With respect to the fixed effects, the coefficients of the significant main effects are given in table 5.18, while the significant interaction effects are shown in table 5.19. As the model predicts log-odds for preposition stranding, a positive *b* coefficient means that a level change of a categorical predictor makes preposition stranding more likely, while a negative *b* estimate reflects a decreasing probability of stranding.

The only predictors which are not involved in interactions are PHRASE_TYPE, FILLER_LENGTH and FREQUENCY_FILLER. The coefficients in table 5.18 show that preposition stranding is negatively associated with PPs contained in NPs, whereas there is no significant difference between VP- and AdjP-contained PPs. As the interaction of VARIETY_TYPE and PHRASE_TYPE did not emerge as significant, this effect of PHRASE_TYPE can be said to affect speakers of all Dynamic Model stages. Since NP-embedded PPs, in which the filler cannot be integrated upon parsing the main verb, are harder to process than PPs contained in VPs or AdjPs, this suggests that speakers of all variety phases have a preference for constructions that involve a reduction in processing cost. Furthermore, these results confirm Hoffmann's (2011: 88) finding that AdjP-contained PPs pattern with NP-contained PPs, which may be due to the fact that AdjP-embedded PPs frequently consist of semantically weak copular verbs such as *be* (cf. 3.1.1).

In addition, increasing filler length leads to a decreasing probability of preposition stranding. On the one hand, this could be due to the fact that in constructions with long fillers such as the ten-word phrase in (5.16) more material needs to be parsed until the filler can be matched with the gap. Moreover, complex filler phrases require more semantic processing than simple fillers such as *which* or *what*.

(5.16) *please state to **which person or particular persons in such company or companies** such payments were so made* <ICE-SIN:S2A-070#57:1:A>

On the other hand, the pied-piping preference of constructions with long fillers may also be the result of an interaction effect of FILLER_LENGTH and PP_TYPE, which could not be tested in the logistic regression analysis. Many pied-piped interrogative constructions with long filler phrases involve adjunct PPs such as *to what extent* (5.17) or *at what time* (5.18). Consequently, the pied-piping preference of constructions with long filler phrases may also be explained by their association with adjunct PPs (also see 5.3.2).

(5.17) ***To what extent** does it affect voters' decision in the electoral processes* <ICE-HK:S2B-040#85:2:A>

(5.18) *I asked you <,,> **at what time** did you visited the accuses house <,,> ?* <ICE-IND:S1B-065#182:1:A>

Furthermore, across all variety phases, stranded constructions are slightly more likely to be used with more frequent fillers. This effect may also reflect a processing principle. More frequent fillers being more deeply entrenched, they have a higher chance

of being activated than less frequent fillers. As a result, high-frequency fillers may be more likely to occur in stranded constructions, in which the filler appears clause-initially. At the same time, the effect of FREQUENCY_FILLER may also be attributed to the frequent occurrence of *what*, which is the most frequent filler in most varieties (cf. figure 5.7; the raw filler frequencies are given in table A.9 in the Appendix), in free RCs and interrogatives, i.e. in clause types that strongly favor stranding (cf. figure 5.1). However, a thorough investigation of this potential interaction effect of FREQUENCY_FILLER and CLAUSE_TYPE requires a larger database.

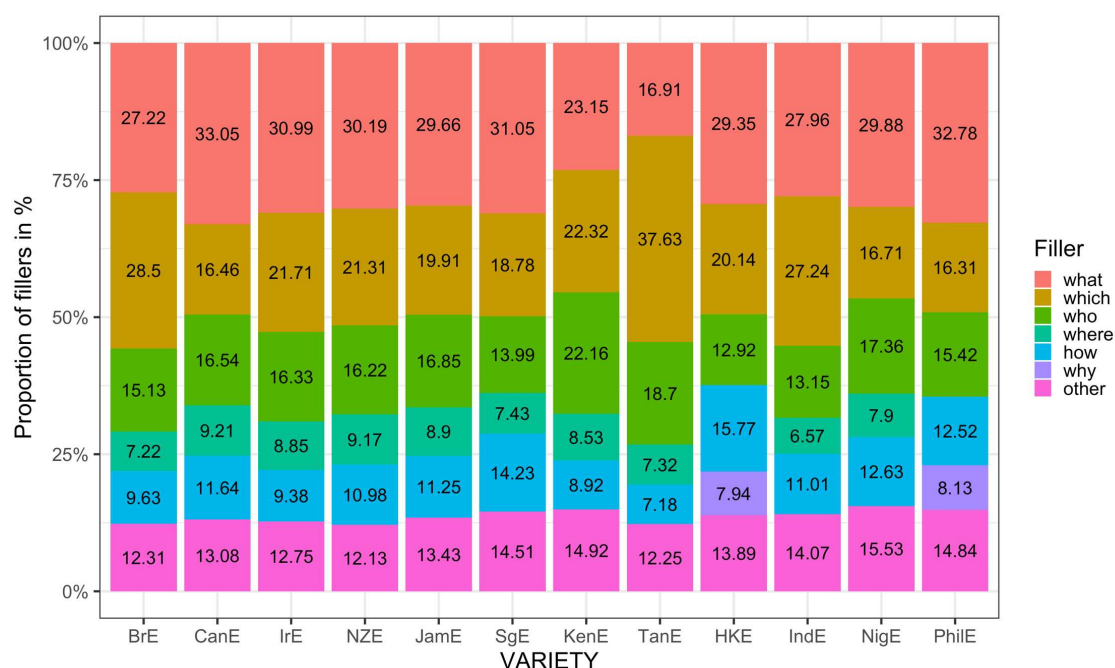


Figure 5.7: Relative filler frequencies per ICE variety

Note that, in contrast to FREQUENCY_FILLER, the variable FREQUENCY_PREPOSITION did not emerge as significant in the logistic regression analysis. Hence, the study does not provide support for the claim that more frequent prepositions are more likely to be stranded than infrequent ones (cf. 3.1.2).

All other significant predictors are involved in interactions (cf. table 5.19). I will first come to the two interaction effects that can be said to apply to all variety phases (as the corresponding three-way interactions including VARIETY_TYPE did not turn out to be significant). As shown in the effects plot in figure 5.8, which illustrates the effect of the interaction while holding all other predictors at constant values (Fox 2003), prepositions in free RCs and interrogatives are generally more likely to be stranded than prepositions in finite RCs. This effect can be attributed to the high processing complexity of finite RCs, which do not only require processing of the FGD but also of the co-

indexation domain of the antecedent noun and the relativizer (cf. 3.2.1). At the same time, the interaction effect of *CLAUSE_TYPE* and *COMPLEXITY* suggests that the different clause types respond differently to increasing complexity. Although the pied-piping preference in finite RCs increases slightly with increasing complexity, the effect of complexity is much stronger for free RCs and interrogatives. Only in constructions with very long FGDs such as (5.19), in which the FGD consists of 20 words, do the latter clause types favor pied-piping. In accordance with Hawkins' (2004: 31) Minimize Domains principle (cf. 2.4.1.1), speakers of all variety phases thus prefer pied-piping, i.e. the variant that reduces the risk of garden paths, in complex constructions in which the filler must be held in working memory over long distances.

(5.19) *Well Mr Loong can you recall **in which month of nineteen ninety the management committee approved the figure of five thousand put forward by the plaintiff*** <ICE-HK:S1B-064#231:1:A>

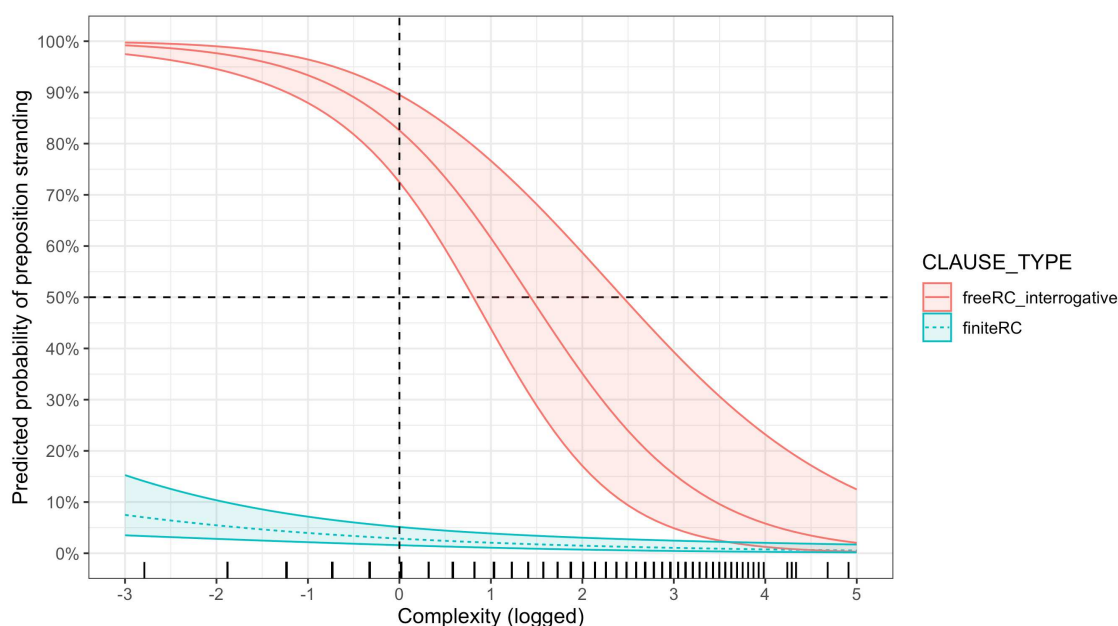


Figure 5.8: The effect of *COMPLEXITY* (logged) * *CLAUSE_TYPE*

Note, however, that the complexity effect of free RCs and interrogatives may, at least in parts, be due to the fact that long interrogative constructions frequently involve adjunct PPs, and in particular respect (5.20a) or degree PPs (5.20b), which themselves strongly favor pied-piping. This interaction effect was, however, too complex to be included in the maximal model.

(5.20) (a) ***On what base** are we to build up the structure of our Indian national life* <ICE-IND:W2B-005#38:1>

- (b) *The fourth question which we will be addressing ourselves to will be **to what extent** are women aware of technical and financial support from donor agencies as well as loan schemes operated by different banks and other organizations in the country <ICE-EA-T:S1B025T>*

Table 5.19: Coefficients of significant interaction effects (of the final model focusing on interactions involving VARIETY_TYPE). Model predictions are for preposition stranding.

Predictors	<i>b</i>	SE	<i>z</i>	<i>p</i>
(Intercept)	6.9625	0.5722	12.1671	<0.001
CLAUSE_TYPE * FORMALITY				
finiteRC + more formal	-1.8507	0.3837	-4.8236	<0.001
CLAUSE_TYPE * COMPLEXITY (logged)				
finiteRC + one-unit increase	0.7489	0.2011	3.7246	<0.001
VARIETY_TYPE * CLAUSE_TYPE				
III + finiteRC	1.8387	0.3427	5.3649	<0.001
IV + finiteRC	1.3271	0.4213	3.1501	<0.01
VARIETY_TYPE * FORMALITY				
III + more formal	1.8094	0.4347	4.1622	<0.001
IV + more formal	0.6844	0.5115	1.3381	0.1809
VARIETY_TYPE * PP_TYPE				
III + complement_like	0.1355	0.3093	0.4381	0.6613
IV + complement_like	0.1365	0.3793	0.3598	0.719
III + adjunct_like	-1.2798	0.4698	-2.7238	<0.01
IV + adjunct_like	-0.372	0.5328	-0.6982	0.485
III + respect_manner_degree	-0.5736	1.2867	-0.4458	0.6558
IV + respect_manner_degree	1.0638	1.1066	0.9613	0.3364

Besides this, the model also identified a significant interaction effect of CLAUSE_TYPE and FORMALITY. Figure 5.9 shows that, regardless of the level of formality, free RCs and interrogatives always prefer stranding, whereas finite RCs always favor pied-piping. Furthermore, across all clause types, stranded prepositions are less likely to be used in more formal contexts than in less formal contexts. At the same time, the significant coefficient in table 5.19 indicates that a change from less formal to more formal contexts makes stranding significantly less likely in finite RCs than in free RCs and interrogatives. This effect is also visualized in figure 5.9, which shows that FORMALITY has a stronger effect on finite RCs than on free RCs and interrogatives. While a change from less formal to more formal finite RCs leads to an almost categorical

pied-piping preference, the probability of preposition stranding is only slightly weaker in formal free RCs and interrogatives than in informal ones. In parts, this finding is in line with Hoffmann (2011: 156), who showed that the pied-piping preference of finite RCs is stronger in formal contexts than in informal usage situations. However, while Hoffmann's analysis suggested that only finite RCs are affected by stylistic variables, the present study indicates that formality also plays a role in free RCs and interrogatives. Importantly, though, the formality effect in these construction types is much weaker than in finite RCs.

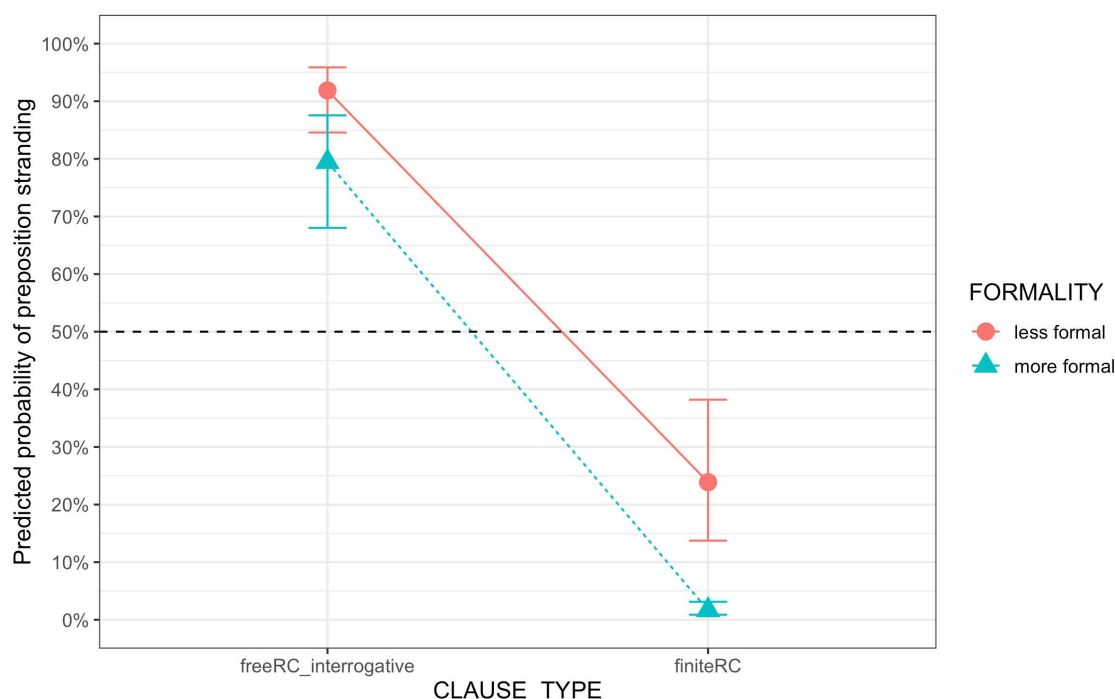


Figure 5.9: The effect of *CLAUSE_TYPE* * *FORMALITY*

In order to fully understand the effect of *FORMALITY*, another interaction must be taken into account, namely that of *VARIETY_TYPE* and *FORMALITY*. Crucially, the significant positive coefficient for Phase III in table 5.19 does not indicate that the stranding probability in formal contexts is higher in Stage III than in Stage V. Instead, this coefficient suggests that, compared to Stage V, Stage III is more likely to use stranded prepositions in more formal contexts than in less formal ones because the overall stranding probability in less formal contexts is lower in Phase III than in Phase V. As illustrated in figure 5.10, the strength of the formality effect is correlated with Dynamic Model stages: While Phases V and IV favor preposition stranding in less formal and pied-piping in more formal contexts, in Phase III the pied-piping preference is only slightly weaker in less formal texts than in more formal ones and pied-piping remains the preferred option regardless of the level of formality. This finding confirms the results of

Hoffmann's (2011: 155, 167) corpus study, which observed a weaker formality effect in KenE than in BrE. Furthermore, these results are in line with previous research which suggested that PCEs are frequently characterized by a lack of register distinctions (cf. 2.3). A cognitive explanation for this effect will be given in section 5.2.4.2, which will take a closer look at the formality effect in finite RCs.

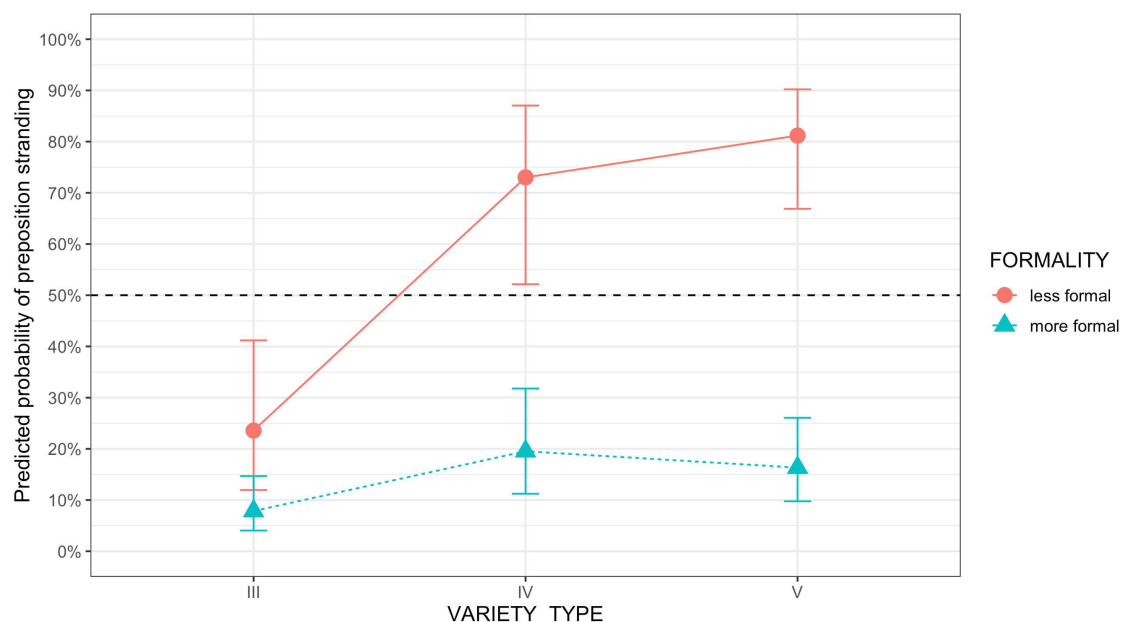


Figure 5.10: The effect of *VARIETY_TYPE* * *FORMALITY*

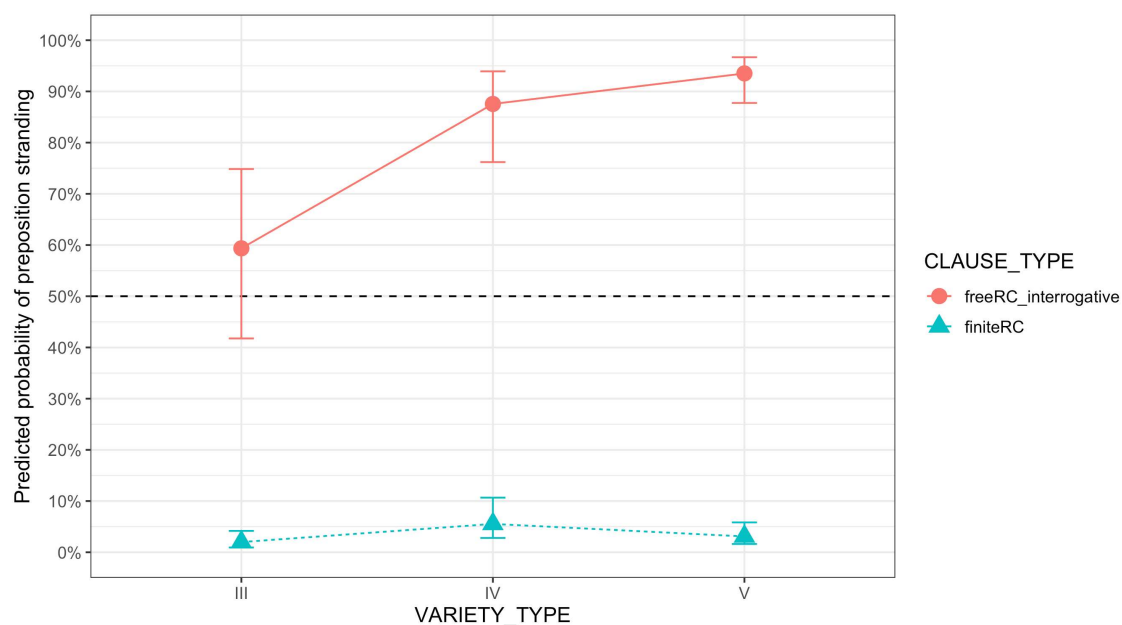


Figure 5.11: The effect of *VARIETY_TYPE* * *CLAUSE_TYPE*

Next, the interaction *VARIETY_TYPE* * *CLAUSE_TYPE* suggests that different variety phases respond differently to *CLAUSE_TYPE*. While speakers of all varieties prefer stranding in the freeRC_interrogative group and pied-piping in finite RCs, figure

5.11 shows that there is a cline from Stage V to Stage III with respect to the strength of the stranding preference in free RCs and interrogatives. While these clause types exhibit an almost categorical stranding preference in the Phase-V varieties, speakers of Phase-III varieties are significantly less likely to use stranded free RC and interrogative constructions. The strong preference for stranded interrogative constructions in Stage V can be explained by the fact that the WH-word in interrogatives should ideally appear clause-initially for discourse-functional reasons. Furthermore, pied-piping in free RCs is normally considered ungrammatical because in free RCs a pied-piped preposition would precede both the antecedent and the relativizer (cf. 3.1.1.2). However, at this point, no explanation can be provided to account for the weaker stranding probability of free RCs and interrogatives in Phase III. Whether this effect is a general tendency among varieties at lower developmental stages or restricted to individual varieties will be examined in section 5.2.3.

In contrast, regarding finite RCs, no clear correlation between variety stages and the strength of the pied-piping preference can be confirmed. Instead, all variety types strongly favor pied-piped finite RC constructions, which can be attributed to the high processing complexity of finite RCs. Surprisingly, Stage IV turned out to favor stranded finite RC constructions slightly more than the other variety types. Since this effect cannot be explained from a processing perspective, section 5.2.3, which focuses on variety-specific effects, will explore whether this weaker pied-piping preference of the Stage-IV varieties can be attributed to cross-linguistic influence.

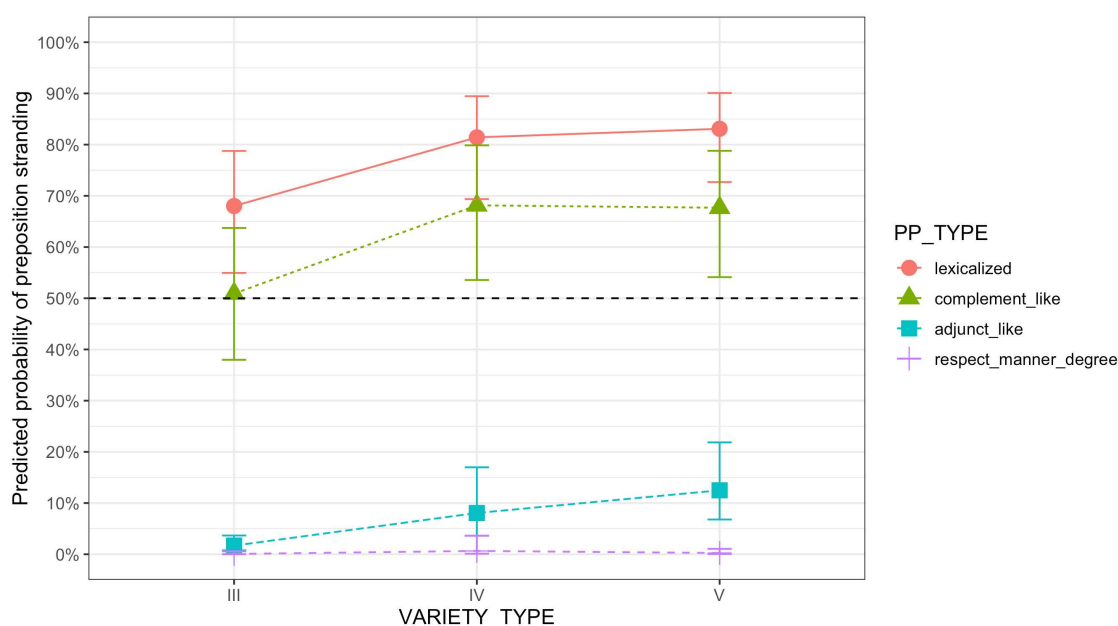


Figure 5.12: The effect of *VARIETY_TYPE* * *PP_TYPE*

The last interaction effect that needs to be addressed is the interaction of `VARIETY_TYPE` and `PP_TYPE`. As shown in figure 5.12, all variety types favor preposition stranding with the groups of lexicalized and complement_like PPs, while pied-piping is preferred with the levels adjunct_like and respect_manner_degree. This shows that all varieties are affected by processing factors: All speakers favor stranding with lexicalized V + P chunks, in which the in-situ placement of the preposition facilitates the interpretation of the verb, while pied-piping is strongly preferred in constructions with respect, manner and degree PPs, in which stranding is usually uninterpretable.

Importantly, however, with respect to preposition placement in adjunct_like PPs, the effects plot again shows a cline from Phase V to Phase III: While Stage III displays an almost categorical pied-piping preference with all adjunct PPs (except movement and accompaniment PPs, which form part of the complement_like group), i.e. with both the adjunct_like and the respect_manner_degree level, adjunct_like PPs are more likely to be stranded in Stage V. Stage IV occupies an intermediate position. In line with the findings of Hoffmann's (2011) magnitude estimation experiments, which observed that KenE exhibits a stronger pied-piping preference with adjunct PPs than BrE (cf. 3.3.1), this effect can be taken as an indication that processing factors have a stronger effect on speakers of less advanced varieties. In constructions with adjunct PPs, pied-piping reduces processing cost as the pied-piped preposition, which provides important information about the integration of the filler, is parsed before the main verb so that garden paths are avoided. Hence, this construction is strongly preferred by speakers of varieties at lower Dynamic Model stages, who receive less input and for whom the use of English involves higher cognitive cost. Considering that Hoffmann (2013: 115) observed a similar tendency among German L2 learners, this almost categorical pied-piping preference with adjunct PPs in the Stage-III varieties might reflect a general processing principle in L2 Englishes. In order to assess to what extent cross-linguistic influence contributes to this effect, the regression model in section 5.2.3 will test the interaction of `PP_TYPE` and `VARIETY`.

Furthermore, figure 5.12 indicates that the stranding preference with lexicalized and complement_like PPs is weaker in Phase III than in Phases IV and V. However, this does not necessarily mean that Stage-III varieties favor stranding with entrenched V + P items less. Instead, recall that the group of lexicalized PPs also includes obligatory complements with *be*, which in Hoffmann's (2011: 240) study emerged as being less strongly associated with stranding in KenE than in BrE (cf. 3.3.1). In order to assess whether the weaker entrenchment of stranded *be*-constructions is a characteristic feature of Stage-III varieties, an HCFA was performed which tested if stranded or pied-piped *be*-

tokens occur more frequently than expected by chance in a specific variety/variety type or clause type.²⁶ Note that an HCFA was conducted to test these associations because the frequencies of the *be*-tokens were not sufficient for the fitting of a regression model (cf. table 5.20). Since the coefficient Q is sample-size independent and the token numbers of the individual varieties are quite low, in the following, only significant types with a Q value of at least 0.005 will be discussed. All significant configurations are provided in table A.10 in the Appendix.

Table 5.20: Stranded and pied-piped be-tokens across varieties (absolute and relative frequencies)

VARIETY (VARIETY_TYPE)	Stranding		Pied-piping		Total	
CanE (V)	121	98.37%	2	1.63%	123	100%
BrE (V)	118	99.16%	1	0.84%	119	100%
IrE (V)	149	99.33%	1	0.67%	150	100%
NZE (V)	147	99.32%	1	0.68%	148	100%
JamE (IV)	136	97.84%	3	2.16%	139	100%
SgE (IV)	89	100.00%	0	0.00%	89	100%
KenE (III)	43	95.56%	2	4.44%	45	100%
TanE (III)	19	79.17%	5	20.83%	24	100%
HKE (III)	49	98.00%	1	2.00%	50	100%
IndE (III)	33	80.49%	8	19.51%	41	100%
NigE (III)	45	90.00%	5	10.00%	50	100%
PhilE (III)	83	92.22%	7	7.78%	90	100%
Column total	1032	96.63%	36	3.37%	1068	100%

As shown in table 5.21, not surprisingly, across all varieties, *be*-tokens occur more frequently than expected by chance in stranded constructions. At the same time, pied-piped *be*-constructions are significantly associated with Phase III and IndE, which suggests that the constraint that requires *be*-constructions to be stranded in the most advanced varieties may be looser in varieties at lower developmental stages. With respect to the logistic regression results, the weaker stranding probability of lexicalized PPs in Phase III may thus be due to the weaker entrenchment of stranded *be*-constructions in these varieties.

²⁶ Note that, in contrast to the logistic regression analysis, this HCFA included all tokens except non-standard types of preposition placement and tokens for which CLAUSE_TYPE could not be coded.

Table 5.21: HCFA results for *VARIETY_TYPE/VARIETY* * *PREPOSITION_PLACEMENT* * *CLAUSE_TYPE* of *be*-tokens (only significant types with $Q \geq 0.005$ are shown)

VARIETY_ TYPE	VARIETY	PREPOSITION_ PLACEMENT	CLAUSE_TYPE	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
.	.	stranding	.	1032	534	464.427	>	1.03595748782809E-254	***	0.933
III	.	pied-piping	.	28	10.1124	31.641	>	1.43361637622153E-05	***	0.017
.	IndE	pied-piping	.	8	1.382	31.6917	>	0.00230086651926	**	0.006
.	.	pied-piping	finite RC	16	0.9438	240.1877	>	7.0632689739992E-14	***	0.014
III	.	pied-piping	finite RC	12	0.2651	519.4563	>	5.57309567454983E-15	***	0.011

A closer look at the pied-piped *be*-tokens in Phase III reveals that most of the filler phrases in these structures are very similar to adjunct PPs. For instance, examples (5.21) - (5.24) were coded as obligatory complements due to the presence of *be*. However, the P + FILLER chunks *on which cot/in what part* in (5.21/5.22) could also be used in constructions with location PPs such as *On which cot did the deceased die?*. The string *until what time* in (5.23) corresponds to a duration/frequency PP chunk, whereas *to what extent* (5.24) is a common structure that normally appears in structures with degree PPs.

(5.21) *Better tell this honourable court <,> in which ward room number <,> and which ward or **on which cot** <,> this deceased was <,> ?* <ICE-IND:S1B-065#58:1:A>

(5.22) *okay **in what part of Germany** are you by the way* <ICE-NIG:ph_04:a645>

(5.23) *Uh **until what time** is your <,> class I mean your next class* <ICE-PHI:S1A-059#58:1:A>

(5.24) *And even then <,> she was having a similar kind of torture <,> torture means I cannot explain you **to what extent** that torture was* <ICE-IND:S1A-069#100:1:B>

Now, considering that adjunct PPs strongly favor pied-piping in Phase III (cf. figure 5.12), speakers of less advanced varieties can be expected to receive a lot of input for pied-piped constructions with adjunct PPs, which should lead to the entrenchment of chunks such as *to what extent* or *in what part*. As these chunks can be activated easily, Phase-III speakers may be likely to rely on them even in constructions with *be*, which strongly favor stranding in Standard varieties. Note, however, that the ICE data also contained 5 pied-piped *be*-tokens from Stage-V varieties. As shown in (5.25) and (5.26), contrary to Johansson & Geisler's (1998: 77) claim that "pied piping is commonly blocked when it would result in a clause-final *be*", stranded *be*-constructions are thus not completely ruled out in Standard varieties either.

(5.25) *The head of the school Father Larkin tells the new teacher, Tom, that they must not disturb the 'ecology' of the society **on whose borders they are*** <ICE-GB:W1B-013 #17:1>

(5.26) *Okay but you do need to know **at what point your break-even sales are*** <ICE-IRE:S2A-028\$A>

The final two significant types identified by the HCFA show that pied-piping with *be* is particularly frequent in finite RCs. Moreover, these pied-piped finite RC tokens with *be* are significantly associated with Phase III. The use of some of these pied-piped finite

RCs tokens can also be attributed to the activation of entrenched chunks such as *way in which* (5.27) or *situation in which* (5.28). In addition, it should be kept in mind that pied-piping is the prototypical option in finite RCs. Since speakers of all varieties should thus have an entrenched pied-piped finite RC construction (see figure 5.11), speakers of less advanced varieties may use this entrenched construction, which can be activated without much effort, regardless of the presence or absence of *be*.

(5.27) *Which is **the third way in which** <,> the rain water is* <ICE-IND:S1B-007#19:1:A>

(5.28) *Speaker Jonathan Ng'eno noticed **the tricky situation in which** Mr Onyancha was and decided o end his agony by ordering him to carry out further investigations into the allegations made by the MPs* <ICE-EA-K:W2E014K>

Returning to the results of the logistic regression analysis, the weaker stranding preference of Phase III with lexicalized PPs can thus be explained by the weaker entrenchment of stranded *be*-constructions in varieties at earlier Dynamic Model stages. At the same time, figure 5.12 shows that complement_like PPs are also less likely to be stranded in varieties at lower stages of evolution. While this effect, at first sight, seems to contradict the assumption that stranded prepositions facilitate processing in the case of entrenched verb-preposition strings, it is important to remember that the regression results may be skewed by characteristics of the data itself. As pointed out in 4.1, ICE-EA, for example, primarily contains formal language, which may lead to a stronger preference for pied-piping in this variety even with complement_like PPs. The logistic regression analysis in 5.2.3 will thus investigate whether the weaker preference for stranding with complement_like PPs in Phase III is due to the data of a single variety.

5.2.2.3 Summary

With respect to hypotheses I-III, the model results can be summarized as follows:

Hypothesis I was confirmed because all variety phases show a preference for constructions that are easy to process. Across all variety types, pied-piping is preferred with NP-embedded PPs, with finite RCs, with adjunct PPs, and in constructions with long fillers or long FGDs. At the same time, stranding is favored with lexically entrenched V + P chunks.

Hypothesis II was confirmed as the strength of the pied-piping preference with adjunct_like PPs turned out to be correlated with Dynamic Model stages, with

Phase III displaying the strongest and Phase V exhibiting the weakest pied-piping preference. This interaction effect of VARIETY_TYPE and PP_TYPE provides support for the claim that processing factors have a stronger effect on less advanced varieties.

Hypothesis III was confirmed because more advanced varieties were shown to exhibit stronger formality effects than varieties at lower stages. While Stages IV and V favor preposition stranding in less formal and pied-piping in more formal contexts, Phase III prefers pied-piping regardless of the level of formality.

While this analysis tested interactions of processing-related factors and stylistic variables with VARIETY_TYPE, in the next section a binomial mixed effects model will be presented which focuses on interactions of processing-related variables and VARIETY.

5.2.3 Preposition placement across varieties

5.2.3.1 Model selection

Table 5.22: Stranded and pied-piped respect, manner, degree PPs across varieties (absolute and relative frequencies)

VARIETY (VARIETY_TYPE)	Stranding		Pied-piping		Total	
CanE (V)	1	0.88%	112	99.12%	113	100%
BrE (V)	0	0.00%	262	100%	262	100%
IrE (V)	2	1.12%	176	98.88%	178	100%
NZE (V)	2	1.09%	182	98.91%	184	100%
JamE (IV)	2	0.95%	208	99.05%	210	100%
SgE (IV)	0	0.00%	92	100%	92	100%
KenE (III)	0	0.00%	173	100%	173	100%
TanE (III)	0	0.00%	148	100%	148	100%
HKE (III)	0	0.00%	133	100%	133	100%
IndE (III)	0	0.00%	224	100%	224	100%
NigE (III)	1	0.82%	121	99.18%	122	100%
PhilE (III)	0	0.00%	114	100%	114	100%
Column total	8	0.41%	1,945	99.59%	1,953	100%

The present logistic regression analysis aims to shed more light on variety-specific effects and manifestations of cross-linguistic influence. However, once more, several variables

had to be recoded or excluded from the maximal model due to multicollinearity and data sparsity issues.

To test the interaction `PP_TYPE * VARIETY`, it was again necessary to conflate several factor levels of the variable `PP_TYPE`. In order to reduce collinearity and to avoid low-frequency cells, the same PP type classification was adopted as in the model focusing on `VARIETY_TYPE` (cf. table 5.9). However, as it turned out that only five varieties have stranded `respect_manner_degree` tokens (cf. table 5.22), all 1,953 `respect`, `manner` and `degree` PPs ultimately had to be eliminated. As these PPs are known to exhibit a much stronger pied-piping preference than the other adjunct PPs (cf. figure 5.12), it was not an option to conflate all adjunct PPs into a single level.

Moreover, due to the low token numbers of pied-piped interrogatives and free RCs, the factor `CLAUSE_TYPE` was again reduced to a two-level distinction between “`freeRC_interrogative`” and “`finite RC`”.

Table 5.23: Tokens with `FILLER_LENGTH` ≥ 3 across varieties (absolute and relative frequencies)

VARIETY (VARIETY_TYPE)	Abs. freq.	Rel. freq
CanE (V)	21	8.86%
BrE (V)	23	9.70%
IrE (V)	18	7.59%
NZE (V)	18	7.59%
JamE (IV)	35	14.77%
SgE (IV)	34	14.35%
KenE (III)	15	6.33%
TanE (III)	6	2.53%
HKE (III)	19	8.02%
IndE (III)	17	7.17%
NigE (III)	9	3.80%
PhilE (III)	22	9.28%
Total	237	100%

Besides this, just like in 5.2.2, the variables `FILLER_TYPE` and `VOICE` were excluded because they are correlated with `CLAUSE_TYPE`. In addition, the token numbers of the `complex_zero` and the `complex_that` levels of `COMPLEXITY_BRIDGE` were again too low to be investigated in the logistic regression analysis. Furthermore, two numeric predictors, namely `FILLER_LENGTH` and `COMPLEXITY`, had to be excluded to avoid overfitting. In section 5.2.2.1, it was pointed out that the dataset contains only

very few observations with complex filler phrases per variety phase. The individual varieties having even lower token numbers of long fillers (cf. table 5.23), testing the interaction of FILLER_LENGTH and VARIETY would not have produced reliable results. At the same time, it was decided not to include FILLER_LENGTH as a main effect since the model aims to detect variety-specific effects. In order to avoid model estimates that are too closely aligned to the idiosyncrasies of the data, FILLER_LENGTH, which had already been tested in the previous analysis (cf. 5.2.2.), was thus excluded altogether.

In addition, the predictor COMPLEXITY was removed from the maximal model as the correlation between COMPLEXITY and CLAUSE_TYPE (cf. figure A.6 in the Appendix) would have required including the three-way interaction COMPLEXITY * CLAUSE_TYPE * VARIETY, which turned out to be too complex for the low token numbers of the individual varieties. Moreover, at least in parts, COMPLEXITY is also encoded by the variable PP_TYPE since constructions with adjunct PPs tend to be longer than constructions with complement PPs. Excluding COMPLEXITY thus also helped to reduce collinearity.

The remaining numeric predictors FREQUENCY_FILLER and FREQUENCY_PREPOSITION were log-transformed, centered and standardized. Note, however, that, in contrast to the model in 5.2.2, the three-way interactions FREQUENCY_FILLER/FREQUENCY_PREPOSITION * FORMALITY * VARIETY had to be excluded to avoid overfitting.

A major difference between the present model and the model outlined in 5.2.2 concerns the variable FORMALITY. Formality having the strongest effect on finite RCs, ideally, the logistic regression analysis would have tested the three-way interaction of VARIETY, FORMALITY and CLAUSE_TYPE. However, this turned out to be impossible due to data sparsity (cf. table A.11 in the Appendix). For instance, the ICE-HK data only contain 3 finite RC tokens from less formal texts. A model estimating parameters for a complex three-way interaction based on such low token numbers would suffer from very high VIFs and extremely large standard errors. As this would increase the Type II error rate, any conclusions drawn from such a model would be unreliable. At the same time, testing the two-way interaction FORMALITY * VARIETY would not have produced reliable results either. As shown in figure 5.13, most informal tokens in the dataset are free RCs and interrogatives, while most tokens from more formal texts are finite RCs. Consequently, the results of a model testing the two-way interaction FORMALITY * VARIETY would be skewed by the correlation of FORMALITY and

CLAUSE_TYPE. Although it would have been interesting to identify variety-specific tendencies regarding stylistic variation, the variable FORMALITY therefore had to be excluded from the model.

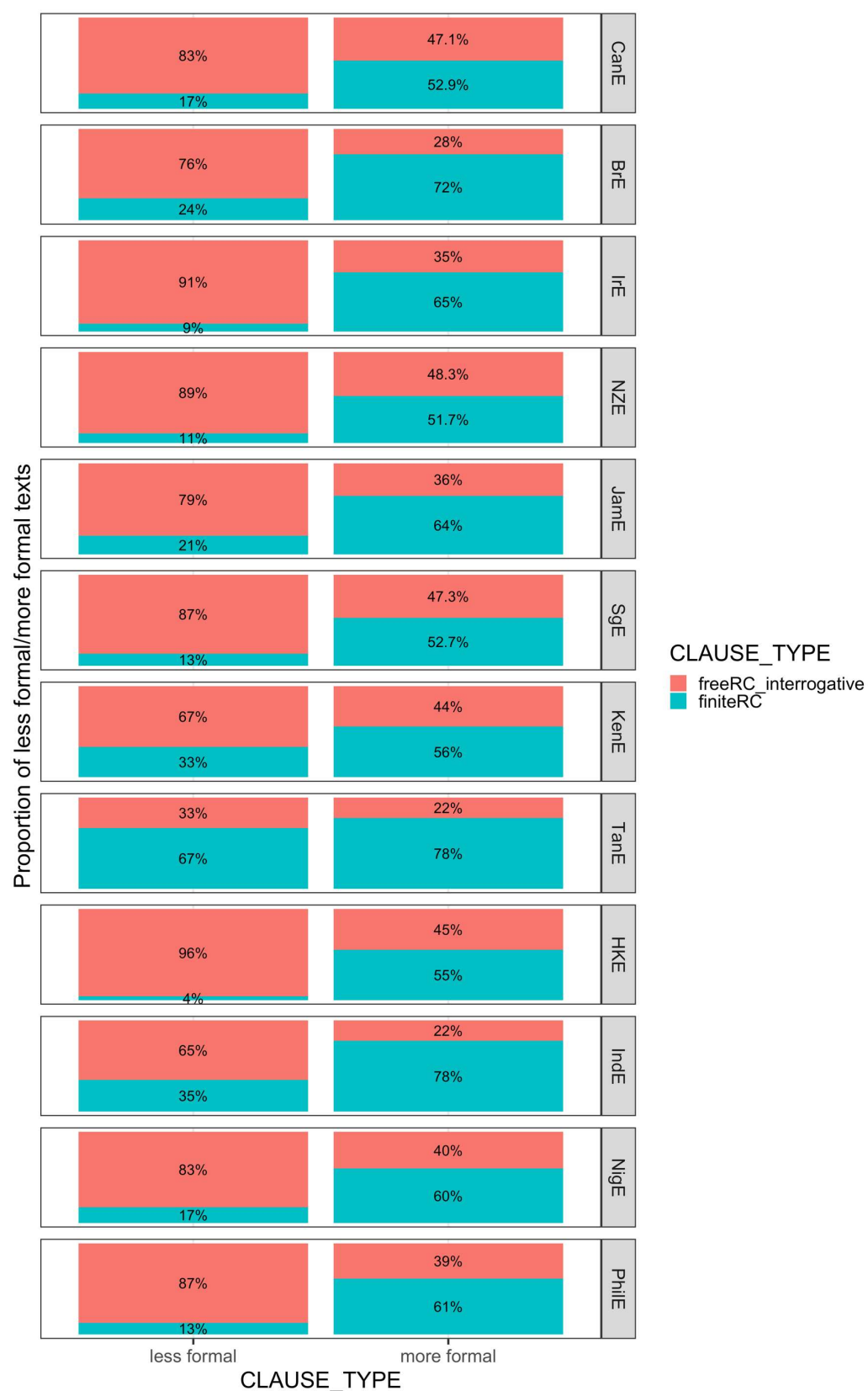


Figure 5.13: Distribution of CLAUSE_TYPE * FORMALITY across VARIETY

Table 5.24: Distribution of PHRASE_TYPE across VARIETY (absolute and relative frequencies)

VARIETY (VARIETY_TYPE)	AdjP		VP		NP		Total	
CanE (V)	21	3.70%	532	93.83%	14	2.47%	567	100%
BrE (V)	35	4.17%	757	90.12%	48	5.71%	840	100%
IrE (V)	27	4.13%	595	91.12%	31	4.75%	653	100%
NZE (V)	31	4.63%	608	90.75%	31	4.63%	670	100%
JamE (IV)	38	4.53%	769	91.77%	31	3.70%	838	100%
SgE (IV)	35	6.53%	477	88.99%	24	4.48%	536	100%
KenE (III)	19	3.70%	473	92.20%	21	4.09%	513	100%
TanE (III)	16	5.30%	259	85.76%	27	8.94%	302	100%
HKE (III)	24	6.43%	338	90.62%	11	2.95%	373	100%
IndE (III)	34	5.55%	523	85.32%	56	9.14%	613	100%
NigE (III)	16	2.91%	513	93.44%	20	3.64%	549	100%
PhilE (III)	39	6.88%	492	86.77%	36	6.35%	567	100%
Column total	335	4.77%	6336	90.24%	350	4.99%	7,021	100%

As the present logistic regression analysis aims to detect variety-specific effects, the maximal model included all interactions of VARIETY and the remaining predictors, i.e. CLAUSE_TYPE, PP_TYPE, FREQUENCY_FILLER and FREQUENCY_PREPOSITION. The only exception is PHRASE_TYPE. Due to very low token numbers of the NP and the AdjP levels (cf. table 5.24), the interaction effect of VARIETY and PHRASE_TYPE could not be tested. While it would have been possible to conflate the AdjP level and the VP level (as these tokens have very similar preposition placement preference), it was not an option to combine NP- and AdjP-contained PPs as the pied-piping preference of NP-embedded PPs is stronger than that of AdjP-embedded PPs (cf. 5.2.2.2). However, since the interaction VARIETY_TYPE * PHRASE_TYPE had not emerged as significant in the previous regression analysis (cf. 5.2.2), PHRASE_TYPE was simply included as a main effect.

While it would have been helpful to explore more three-way interactions such as CLAUSE_TYPE * PP_TYPE * VARIETY, this was, once again, not possible due to data sparseness. With respect to the interpretation of the results for CLAUSE_TYPE, it should thus be kept in mind that the results might be skewed since the model does not include parameters for potential interaction effects of CLAUSE_TYPE, FORMALITY, and PP_TYPE.

Coming to the random effects, random intercepts were included for FILE_ID and PREPOSITION. Again, all prepositions for which the data contain less than 5 tokens ($N = 38$) were combined into one group “other”, whereas the levels of FILE_ID with only one observation ($N = 1,694$) were not pooled together as it was not considered reasonable to conflate tokens from 1,694 individual speakers into a single level. Due to convergence problems, no variance terms were included for random slopes or correlations between random effects.

After all nonsignificant random and fixed effects had been removed from the maximal model with the help of LRTs, the model structure given in (5.29) emerged as the final model:

$$(5.29) \text{ PREPOSITION\_PLACEMENT} \sim \\ \text{VARIETY} + \text{CLAUSE\_TYPE} + \text{PP\_TYPE} + \\ \text{FREQUENCY\_FILLER} + \text{PHRASE\_TYPE} + \\ \text{VARIETY:CLAUSE\_TYPE} + \\ \text{VARIETY:PP\_TYPE} + \\ (1 \mid \text{FILE\_ID}) + (1 \mid \text{PREP})$$

The final model is highly significantly better ($p < 0.001$) than a null model which includes the same random effects structure but no fixed effects. The marginal $R^2 = 0.68$ and the conditional $R^2 = 0.89$. Collinearity should not be an issue as all adjusted GVIFs are below 3, and the condition number K is 22.94, i.e. clearly below 30, which would indicate harmful collinearity. Furthermore, the classification accuracy of the model is 96.14% (precision = 95.39%, recall = 96.40%), which is highly significantly better than a model that always predicts pied-piping ($p < 0.001$). That the model has excellent predictive capacities is also suggested by the C -index of 0.99.

5.2.3.2 Results

The random intercept estimates for FILE_ID and PREPOSITION are given in table 5.25, while the intercept adjustments for the various prepositions are visualized in figure 5.14. Note that positive intercept adjustments again indicate a stranding preference, whereas negative values indicate a pied-piping preference. Not surprisingly, the results are very similar to those of the model described in section 5.2.2.2. *Among*, *around* and *since* are the prepositions which are most likely to be pied-piped, whereas *against*, *through* and *about* have the highest stranding probability. Now, in 5.2.2.2, I suggested that the pied-piping preference of *around* and *among* might be due to the fact that these prepositions

frequently occur in respect PPs, i.e. in PP types which have an almost categorical pied-piping preference. However, since all respect, manner and degree PPs were excluded from the present model, the strong pied-piping preference of *around* and *among* may indeed be due to idiosyncratic lexical effects of these prepositions.

Table 5.25: Variance and standard deviation estimates of random effects (of the final model focusing on interactions involving VARIETY)

Groups	Variance	SD
FILE_ID	4.548	2.133
PREPOSITION	1.49	1.221

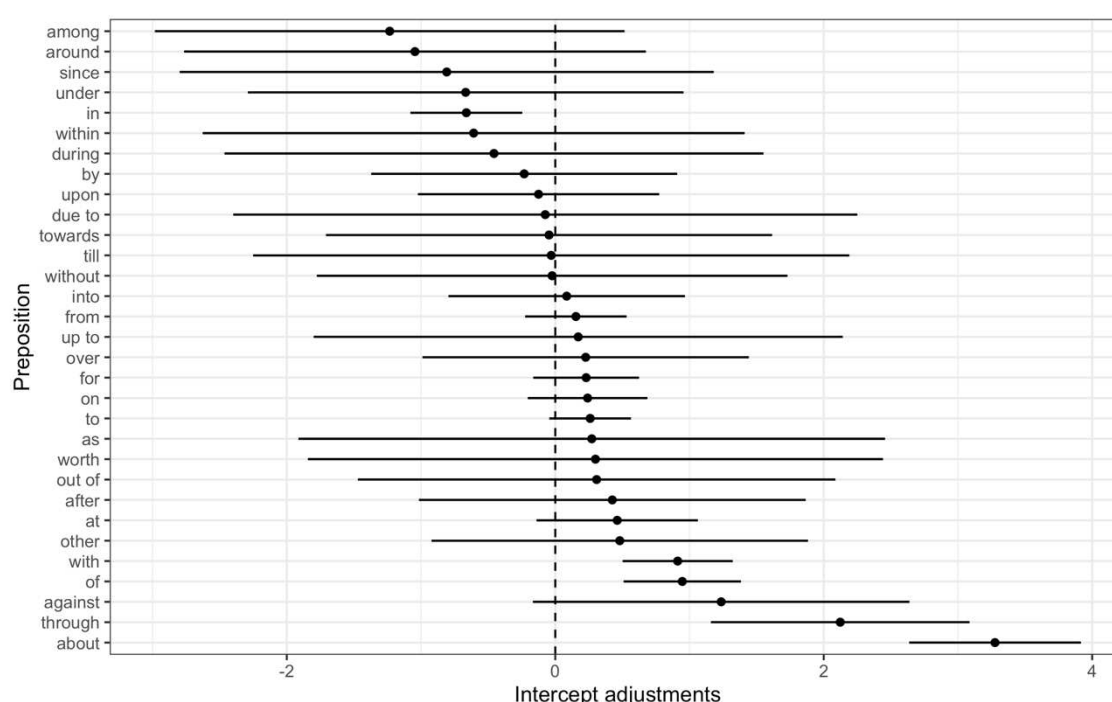


Figure 5.14: Intercept adjustments of PREPOSITION and Wald confidence intervals (of the final model focusing on interactions involving VARIETY)

Turning to the fixed effects, the estimates of the significant main effects are listed in table 5.26, whereas the coefficients of the significant interaction effects are given in table 5.27. The model predicts again log-odds for preposition stranding, which means that a positive b value indicates that a level change of a categorical predictor goes along with an increasing probability of preposition stranding, whereas a negative b value indicates that preposition stranding becomes less likely.

Beginning with the predictors that are not involved in interactions, the model suggests that there is no significant difference between PPs embedded in VPs and AdjPs, while preposition stranding is significantly less likely with NP-contained PPs. The model thus confirms the claim made in 5.2.2.2 that stranded constructions with NP-embedded

PPs are generally avoided since these constructions do not allow speakers to match the filler with the gap upon encountering the main verb. Furthermore, the results again suggest that the probability of stranding increases slightly with more frequent fillers. As pointed out in 5.2.2.2, more frequent and more deeply entrenched fillers, which have a high chance of being activated, may favor stranding because stranded constructions ensure that the filler appears clause-initially. However, remember that this effect of FILLER_FREQUENCY may also be due to an interaction of FREQUENCY_FILLER and CLAUSE_TYPE, which could not be tested in the model. As indicated in 5.2.2.2, for instance, *what* is the most frequent filler in most varieties but also strongly associated with free RCs and interrogatives, i.e. clause types with a very strong stranding preference.

Table 5.26: Coefficients of significant main effects (of the final model focusing on interactions involving VARIETY). Model predictions are for preposition stranding.

Predictors	<i>b</i>	SE	<i>z</i>	<i>p</i>
(Intercept)	5.3771	0.6484	8.2926	<0.001
VARIETY				
BrE > CanE	0.8574	0.9188	0.9332	0.3507
BrE > IrE	0.3718	0.7793	0.4771	0.6333
BrE > NZE	1.0514	0.8811	1.1933	0.2327
BrE > JamE	-0.7533	0.6877	-1.0954	0.2734
BrE > SgE	0.4756	0.8231	0.5778	0.5634
BrE > KenE	-1.2914	0.7386	-1.7484	0.0804
BrE > TanE	-1.1548	0.9338	-1.2367	0.2162
BrE > HKE	0.2392	0.9515	0.2514	0.8015
BrE > IndE	-3.2308	0.708	-4.563	<0.001
BrE > NigE	-1.4931	0.699	-2.1361	<0.05
BrE > PhilE	-1.8623	0.6911	-2.6945	<0.01
CLAUSE_TYPE				
freeRC_interrogative > finiteRC	-6.9853	0.6223	-11.2252	<0.001
PP_TYPE				
lexicalized > complement_like	-1.0245	0.4411	-2.3229	<0.05
lexicalized > adjunct_like	-3.4892	0.5745	-6.0732	<0.001
FREQUENCY_FILLER (logged)				
one-unit increase	0.4462	0.0639	6.9867	<0.001
PHRASE_TYPE				
VP > AdjP	-0.1203	0.261	-0.461	0.6448
VP > NP	-1.7208	0.3195	-5.3865	<0.001

Table 5.27: Coefficients of significant interaction effects (of the final model focusing on interactions involving *VARIETY*). Model predictions are for preposition stranding.

Predictors	<i>b</i>	SE	<i>z</i>	<i>p</i>
VARIETY * CLAUSE_TYPE				
CanE + finiteRC	-0.3806	0.8391	-0.4536	0.6501
IrE + finiteRC	-0.0753	0.7135	-0.1055	0.916
NZE + finiteRC	0.4431	0.8031	0.5517	0.5812
JamE + finiteRC	1.0678	0.6451	1.6554	0.0979
SgE + finiteRC	2.0911	0.7273	2.8753	<0.01
KenE + finiteRC	1.9139	0.6765	2.8293	<0.01
TanE + finiteRC	1.8325	0.8633	2.1227	<0.05
HKE + finiteRC	2.1088	0.8108	2.6008	<0.01
IndE + finiteRC	2.1366	0.6779	3.152	<0.01
NigE + finiteRC	2.3325	0.6834	3.4132	<0.001
Phile + finiteRC	2.5423	0.6854	3.7091	<0.001
VARIETY * PP_TYPE				
CanE + complement_like	0.4247	0.7974	0.5326	0.5943
IrE + complement_like	0.2675	0.6593	0.4057	0.685
NZE + complement_like	0.3706	0.6793	0.5457	0.5853
JamE + complement_like	0.8651	0.5781	1.4966	0.1345
SgE + complement_like	-0.86	0.7	-1.2285	0.2193
KenE + complement_like	0.3319	0.6538	0.5076	0.6117
TanE + complement_like	-1.1552	0.7474	-1.5456	0.1222
HKE + complement_like	-1.7581	0.8526	-2.062	<0.05
IndE + complement_like	0.2693	0.6338	0.4249	0.6709
NigE + complement_like	1.3986	0.6314	2.215	<0.05
Phile + complement_like	1.0635	0.6311	1.6851	0.092
CanE + adjunct_like	0.2144	0.9106	0.2355	0.8138
IrE + adjunct_like	0.1225	0.8125	0.1508	0.8801
NZE + adjunct_like	-1.3322	0.9161	-1.4541	0.1459
JamE + adjunct_like	-0.4993	0.7831	-0.6376	0.5237
SgE + adjunct_like	-1.99	0.9291	-2.1418	<0.05
KenE + adjunct_like	-2.4703	0.9891	-2.4976	<0.05
TanE + adjunct_like	-2.6053	1.2255	-2.1259	<0.05
HKE + adjunct_like	-4.2234	1.2621	-3.3464	<0.001
IndE + adjunct_like	-2.9913	1.1379	-2.6289	<0.01
NigE + adjunct_like	-1.2637	0.9291	-1.36	0.1738
Phile + adjunct_like	-1.2264	0.9644	-1.2717	0.2035

All other predictors are involved in interactions, which indicates that there are variety-specific effects regarding `CLAUSE_TYPE` and `PP_TYPE`. Starting with `PP_TYPE`, the model corroborates the prediction that all speakers are strongly guided by processing constraints because, apart from IndE, all varieties prefer stranding with lexicalized and complement_like PPs, whereas pied-piping is the more likely choice with adjunct_like PPs. At the same time, the model confirms the correlation between Dynamic Model stages and the strength of the pied-piping preference with adjunct_like PPs which was observed in 5.2.2.2: The effects plot in figure 5.15 clearly shows that all Stage-III varieties (KenE, TanE, HKE, IndE, NigE and PhlE) display an almost categorical pied-piping preference with these PP types, while the Stage-V varieties (particularly CanE and IrE) are most likely to use stranding constructions. The Stage-IV varieties JamE and SgE, once again, show a weaker pied-piping preference than the Stage-III varieties but a stronger pied-piping preference than the Stage-V varieties. Note that this effect is unlikely to be due to cross-linguistic influence as it applies to varieties with a wide range of typologically distinct L1s. Moreover, for instance, in the local languages of Nigeria, both stranded and pied-piped constructions exist, while Jamaican Creole only permits stranding (cf. 3.4). Taking into account that Hoffmann (2011, 2013) observed similar effects in L2 Englishes, the almost categorical pied-piping preference with adjunct PPs in the less advanced varieties may thus indeed be attributed to the fact that the constructional choices of varieties at earlier stages are more strongly driven by processing constraints than those of more advanced varieties. Such an explanation is consistent with studies on *that*-complementation (e.g. Wulff, Lester & Martinez-Garcia 2014; Wulff, Gries & Lester 2018) or particle placement (e.g. Gilquin 2015: 64–65; Wulff & Gries 2019) which also found that L2 varieties favor structures that increase parsing efficiency (cf. 2.4.1), i.e. constructions which reduce the risk of garden paths and the number of elements that need to be processed (Hawkins 2004: 25, 53).

Furthermore, in line with the results of the previous regression analysis (cf. 5.2.2.2) which suggested that lexicalized PPs have a lower stranding probability in varieties at earlier stages of evolution, JamE, KenE, TanE, IndE, NigE and PhlE emerged as the varieties which are least likely to use stranded prepositions with lexicalized PPs. This effect can again be explained by the weaker entrenchment of stranded *be*-constructions in Phase III (cf. 5.2.2.2). The only clear exception is IndE, which prefers pied-piping across all PP type levels (though stranding is also most likely with lexically entrenched *V + P* combinations). Importantly, a change from lexicalized to complement_like PPs makes stranding slightly more likely in IndE compared to BrE (cf.

the positive coefficient for IndE complement_like PPs in table 5.27) not because IndE has a stronger stranding preference with complement_like PPs but rather because lexicalized PPs are less likely to be stranded in IndE than in BrE. A potential explanation for the strong pied-piping preference of IndE will be given at the end of this section.

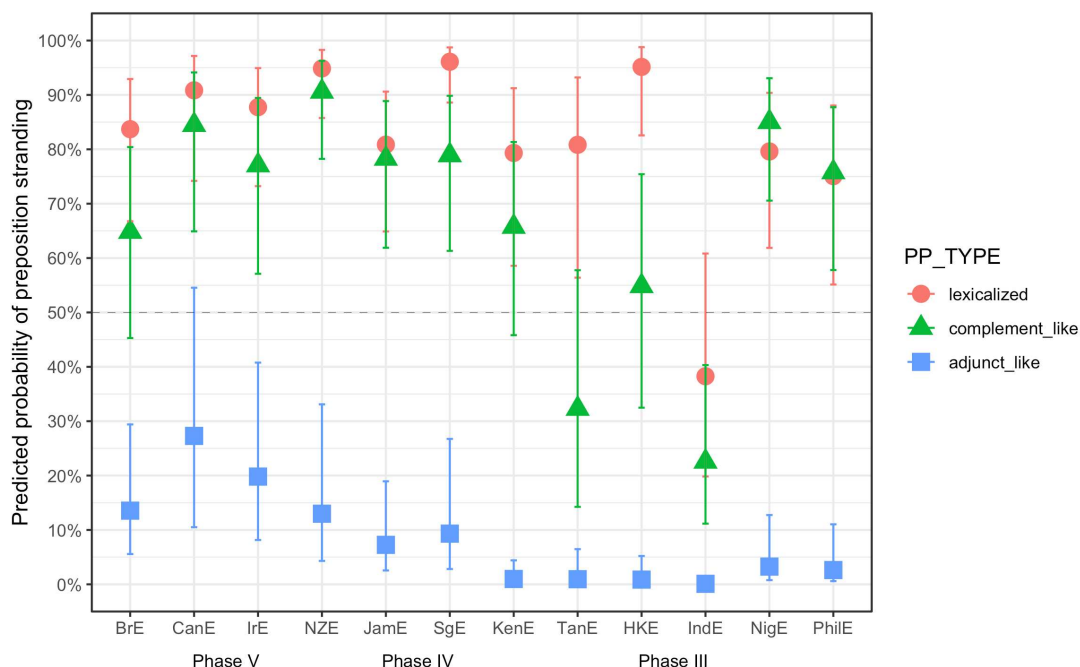


Figure 5.15: The effect of *VARIETY * PP_TYPE*

Moreover, recall that the model in 5.2.2.2 indicated that Phase III has a weaker stranding preference with complement_like PPs than higher Dynamic Model stages. The present logistic regression analysis shows that this effect cannot be attributed to the data of a single variety because, apart from IndE, TanE and HKE also have a relatively strong preference for pied-piping with complement_like PPs, which is slightly surprising since these PPs were assumed to mildly prefer stranding. Note that the effect of TanE cannot be attributed to cross-linguistic influence because the substrate systems of TanE and KenE are very similar, and the stranding preference with complement_like PPs is much stronger in KenE. Instead, the weaker stranding probability of TanE can be explained by the corpus design of ICE-EA-T. As the Tanzanian ICE component hardly includes informal language (cf. 4.1), it may contain a lower proportion of optional complement tokens such as *What do you want to talk about?*, which typically involve stranded prepositions and which appear quite frequently in the private dialogue categories of the other ICE components. The weaker stranding preference of HKE is also unlikely to reflect cross-linguistic influence because speakers of HKE should be subject to similar substrate

influences as speakers of SgE, who are, however, much more likely to use stranded prepositions in constructions with complement_like PPs.

Crucially, the stronger pied-piping preference with complement_like PPs in varieties such as HKE, IndE or TanE does not indicate that speakers of less advanced varieties favor stranding with entrenched V + P chunks less. This can, for instance, be seen in the fact that HKE strongly favors stranding with lexicalized PPs although the stranding preference with complement_like PPs is relatively weak. However, in Hong Kong, for example, the use of English is largely restricted to written contexts and “on almost all occasions Cantonese is the sole language of spoken communication” (Bolton 2012: 229). Now, remember that the group of complement_like PPs includes accompaniment and movement adjuncts, which have been argued to prefer stranding as they frequently involve entrenched V + P chunks (cf. 3.2.2). As a result of the limited amount of input that is characteristic of Phase III, these verb-preposition strings may be less deeply entrenched in the construction networks of speakers of less advanced varieties. At the same time, from a corpus-as-input point of view, figure 5.15 (and figure 5.12) suggests that speakers of varieties at lower Dynamic Model stages encounter adjunct PPs almost exclusively with pied-piped prepositions, which indicates that these speakers receive a lot of positive input for P + WH combinations. Since pied-piping constructions can thus be assumed to be more deeply entrenched in varieties at earlier stages than constructions with stranded prepositions, these speakers may prefer pied-piping even with complement_like PPs which clearly favor stranding in the more advanced varieties. In order to assess this claim, in section 5.3 HCFAAs will be performed to identify entrenched V + P and P + WH combinations in the individual varieties.

The second significant interaction identified by the model, namely that of VARIETY and CLAUSE_TYPE, confirms that speakers of all varieties prefer stranding in free RCs and interrogatives but pied-piping in finite RCs. Only speakers of IndE are almost equally likely to use stranded and pied-piped prepositions in the freeRC_interrogative group. Moreover, just like in the effects plot which visualized the interaction of CLAUSE_TYPE and VARIETY_TYPE (cf. figure 5.11), figure 5.16 shows a correlation between Dynamic Model stages and the strength of the stranding preference in free RCs/interrogatives. While the four Phase-V varieties exhibit an almost categorical stranding preference, the six Phase-III varieties are clearly less likely to use stranded prepositions in these clause types, with the Phase-IV varieties being situated somewhere in-between. This suggests that the effect observed in section 5.2.2.2 is not limited to individual varieties, but that earlier Dynamic Model phases generally have a

stronger preference for pied-piping in free RCs and interrogatives than varieties at higher evolutionary stages.

Table 5.28: Stranding and pied-piping per PP type in WH-interrogatives across VARIETY_TYPE (absolute and relative frequencies)

Phase III	Stranding		Pied-piping		Row total	
idiosyncratic	134	98.53%	2	1.47%	136	100%
lexicalized	274	90.73%	28	9.27%	302	100%
complement_like	336	90.81%	34	9.19%	370	100%
adjunct_like	6	5.83%	97	94.17%	103	100%
respect_manner_degree	0	0%	85	100%	85	100%
Column total	750	75.30%	246	24.70%	996	100%
Phase IV	Stranding		Pied-piping		Row total	
idiosyncratic	110	99.10%	1	0.90%	111	100%
lexicalized	179	98.35%	3	1.65%	182	100%
complement_like	197	93.81%	13	6.19%	210	100%
adjunct_like	23	51.11%	22	48.89%	45	100%
respect_manner_degree	0	0%	19	100%	19	100%
Column total	509	89.77%	58	10.23%	567	100%
Phase V	Stranding		Pied-piping		Row total	
idiosyncratic	222	100%	0	0%	222	100%
lexicalized	367	97.61%	9	2.39%	376	100%
complement_like	344	97.18%	10	2.82%	354	100%
adjunct_like	86	78.18%	24	21.82%	110	100%
respect_manner_degree	2	4.76%	40	95.24%	42	100%
Column total	1,021	92.48%	83	7.52%	1,104	100%

One reason that might favor the use of pied-piped interrogative constructions in the Stage-III varieties is cross-linguistic influence. As outlined in section 3.4.1, all Phase-III varieties have local languages that either have pied-piped or in-situ WH-interrogative constructions. Due to cross-linguistic influence, the discourse-functional constraint that leads to the clause-initial placement of the WH-word in interrogatives in Standard varieties may thus not be as strong in these PCEs. At the same time, it is unlikely that this effect can be solely attributed to cross-linguistic influence because the Nigerian languages Yoruba, Tiv, Hausa and Nigerian Pidgin also license preposition stranding in interrogatives. Although the stranding preference in NigE is slightly stronger than in the other Stage-III varieties, it is still clearly weaker than in the more advanced varieties. In addition, figure 5.16 indicates that the stranding preference in free RCs and interrogatives

is also slightly weaker in JamE than in the Stage-V varieties even though Jamaican Creole only permits preposition stranding. Hence, the stronger preference for pied-piped interrogative constructions in the less advanced varieties cannot be exclusively explained by substrate influence.

Instead, remember that the model also suggested a cline from Phase V to Phase III with respect to the pied-piping preference with adjunct PPs, which is almost categorical in the six Stage-III varieties but weaker in the four Stage-V varieties. Although the three-way interaction `PP_TYPE * CLAUSE_TYPE * VARIETY` could not be tested in the logistic regression analysis, this may suggest that less advanced varieties have a stronger preference for pied-piped adjunct WH-interrogative²⁷ constructions than more advanced varieties. This assumption is supported by the fact that the entire Stage-III data contain only 6 stranded adjunct_like interrogative tokens (cf. table 5.28). In order to test this claim, in section 5.3.2, an HCFA will be conducted which will take a closer look at all `P + FILLER` combinations used in pied-piped interrogatives.

At the same time, just like in the model which focused on effects of `VARIETY_TYPE`, no correlation between variety stages and the strength of the pied-piping preference was found with respect to finite RCs. Instead, SgE, NZE and NigE emerged as the varieties that are most likely to use stranded finite RC constructions, while it is again IndE that exhibits the strongest pied-piping preference. The model therefore confirms the claim that all varieties favor pied-piping due to the high processing cost of finite RCs, which do not only require processing of the FGD but also of the co-indexation domain of the antecedent and the relativizer. However, as PCEs such as NigE, PhilE, HKE, JamE or SgE display a stronger stranding preference than BrE, the strength of this pied-piping preference does not correlate with variety stages and does thus not seem to be linked to processing factors. In order to explain the strong stranding preference of SgE, cross-linguistic influence may be relevant. Remember that prepositions are normally dropped under relativization in Mandarin Chinese, while Malay does not license the relativization of prepositional objects at all (cf. 3.4.2). Furthermore, SgE favors *that*-relatives, in which stranding is obligatory, over *which*-relatives (Suárez-Gómez 2015: 257–258). The combination of both factors, i.e. the absence of pied-piped relative clause constructions in local languages and the frequent exposure to *that*-relatives, which might

²⁷ As the Stage-III data only contains 10 pied-piped adjunct_like free RC tokens, the effect in the `freeRC_interrogative` group is most likely not due to free RCs.

serve as a model for WH-relatives, may lead to an overall preference for stranded finite RC constructions in SgE.

In contrast, the emergence of NigE as the Stage-III variety with the strongest stranding preference cannot be attributed to cross-linguistic influence. Even though Yoruba has stranded relative clause constructions, stranding is prohibited in both Hausa and Igbo. Furthermore, apart from IndE, NigE is the only variety with major L1s, namely Hausa and Yoruba, that have pied-piped relative constructions (cf. 3.4.2). Consequently, it is unlikely that cross-linguistic influence leads to a preference for stranded relative clause constructions in NigE. Likewise, cross-linguistic influence cannot account for the fact that the stranding probability of PhilE is slightly higher than that of KenE, TanE, HKE and IndE since the Philippine languages do not license preposition stranding. The observed differences between varieties may thus also be linked to the data itself as some ICE components may simply contain a larger proportion of finite RCs with PP types that favor stranding than others. Moreover, it is important to remember that the interaction between `CLAUSE_TYPE` and `FORMALITY` could not be tested in the model. For example, varieties with a weaker pied-piping preference such as NZE may have a larger proportion of informal finite RCs, which are more likely to be stranded than formal ones. The interactions of clause type, PP type, formality and variety will thus have to be investigated in future studies using larger corpora.

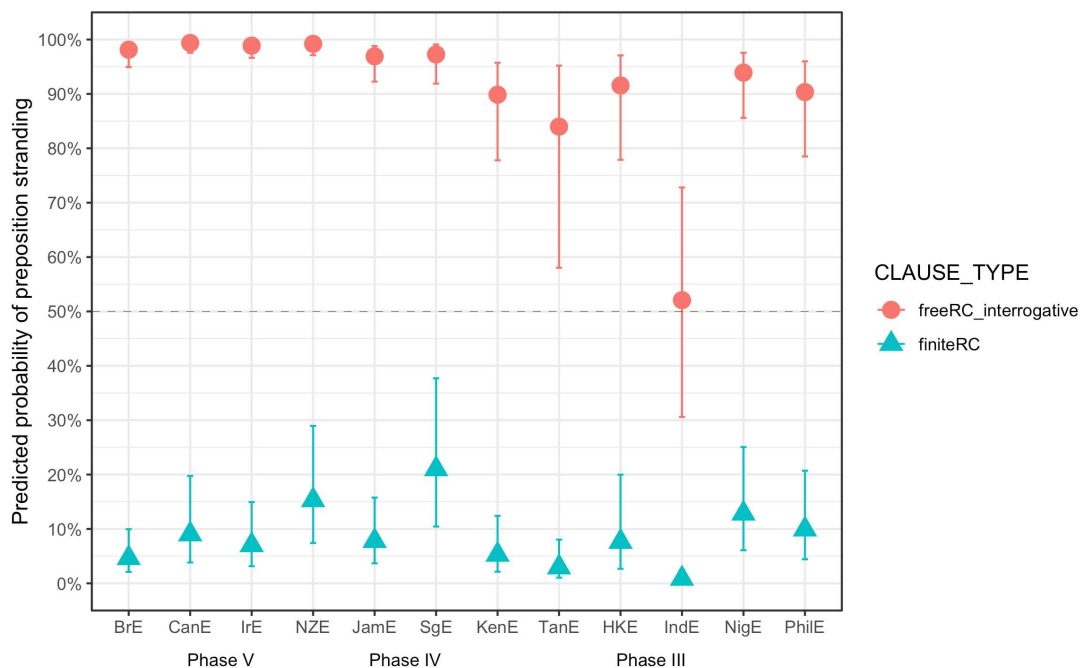


Figure 5.16: The effect of *VARIETY* * *CLAUSE_TYPE*

The last effect that requires an explanation is the strong pied-piping preference of IndE, which stands out from all other varieties. The first reason that comes to mind in order to explain this effect is cross-linguistic influence. As outlined in 3.4, the L1s spoken in India typically have pied-piped or in-situ interrogative constructions. Furthermore, the Indo-Aryan languages, which account for almost 80% of the local languages spoken in India, have relative constructions in which postpositions are pied-piped. Now, as pointed out above, in-situ WH-interrogative constructions, which could contribute to the use of P + WH items in English, exist in a number of L1s spoken in the Stage-III varieties. However, from what I could glean from the literature, apart from NigE, IndE is the only variety with major local languages in which pied-piping is also the norm in relative clauses (cf. table 3.4). Even though these structures involve post- instead of prepositions, cross-linguistic influence might thus lead to a preference for constructions in IndE in which the preposition and the WH-element are not separated. Furthermore, IndE is generally assumed to be characterized by a very formal and conservative style. As Sharma (2017: 325) pointed out, “[t]he prominence of British-influenced school curricula and the high prestige of functional literacy in English [...] perpetuates a characteristic conservatism, formal style, [...] and widespread prescriptivism in Indian English”. Formal registers being associated with pied-piping in BrE, particularly in the context of language teaching, speakers of IndE might thus be exposed to a lot of positive input for pied-piped prepositions. This may contribute to the entrenchment of pied-piping constructions and the overuse of pied-piped prepositions in environments in which stranding would be preferred in Standard varieties.

5.2.3.3 Summary

The results of the present logistic regression analysis, which was conducted in order to identify interaction effects between processing-related factors and the variable VARIETY, confirmed the effects of VARIETY_TYPE found in 5.2.2.2. Particularly concerning preposition placement with interrogatives and adjunct PPs, the similarities between varieties of the same developmental stage are striking. While hypothesis III could not be tested, with respect to hypothesis I, II and VII, the following conclusions can be drawn from the regression model:

Hypothesis I was confirmed because, once again, all varieties turned out to favor prototypical constructions. Pied-piping is generally preferred with NP-contained

PPs, with adjunct PPs, and in finite RCs, whereas stranding is favored in constructions with lexicalized V + P combinations.

Hypothesis II was also confirmed as the model identified a cline from Phase V to Phase III regarding preposition placement with adjunct PPs: While the Stage-III varieties almost categorically use pied-piped prepositions in constructions with adjunct PPs, more advanced varieties are more likely to strand prepositions in the same context. Pied-piped constructions with adjunct PPs being easier to process than stranded ones, this suggests that the constructional choices of speakers of less advanced varieties are more strongly driven by processing constraints than those of varieties at higher stages of evolution.

The claim that cross-linguistic influence should lead to quantitative innovations in PCEs was partly confirmed. The following findings do not provide support for hypothesis VII:

JamE does not exhibit a stronger stranding preference than the other PCEs although JamE is the variety in which speakers should be most likely to encounter stranded (both interrogative and relative clause) constructions in their L1 input.

There is no difference between the Stage-IV varieties SgE and JamE with respect to the stranding preference in the freeRC_interrogative group although the local languages spoken in Singapore do not license stranded interrogative constructions, whereas stranding is the only option in Jamaican Creole.

NigE is the Stage-III variety with the strongest stranding preference in finite RCs although, apart from IndE, NigE is the only New English variety with major local languages that license pied-piping in relative clauses.

At the same time, the following effects can, at least in parts, be attributed to cross-linguistic influence:

The weaker preference for stranded interrogative constructions in HKE, PhilE, KenE and TanE may be favored by the fact that the local languages of these varieties only license pied-piped or in-situ interrogative constructions.

NigE exhibits a slightly stronger stranding preference in interrogative constructions than the other Stage-III varieties, which can be explained

by the fact that NigE is the only Stage-III variety in which speakers also receive input for stranded interrogative constructions.

The strong pied-piping preference of IndE may be attributed to the fact that IndE is the variety in which speakers receive the most input for pied-piping constructions (particularly for pied-piped relative clause constructions).

So far, the results have indicated that a variety of factors, namely processing constraints, input effects, and cross-linguistic influence, lead to differences in constructional preferences between varieties. Finite RCs being more complex than other filler-gap constructions (cf. 3.2.1), the next logistic regression analysis will concentrate on the constructional competition between preposition stranding and pied-piping in finite RCs.

5.2.4 *Preposition placement in finite WH-relative clause constructions*

5.2.4.1 Model selection

In order to understand how processing constraints and formality affect preposition placement in finite RCs, a binomial mixed effects model based on all finite RC tokens was fitted which focused on interactions between VARIETY_TYPE, processing factors and stylistic variables.²⁸ Since this logistic regression analysis only included the finite RC data, several variables were tested which could not be explored in the previous models (cf. 5.2.2 and 5.2.3). Note, however, that the low token numbers of the individual varieties made it impossible to test the factor VARIETY (cf. table 5.29; for the finite RC frequencies per VARIETY_TYPE, see table 5.10).

With respect to processing-related factors, the regression analysis investigated the factors PHRASE_TYPE, PP_TYPE and COMPLEXITY as well as the variable RESTRICTIVENESS, which is only relevant in connection with finite RCs. However, one token for which RESTRICTIVENESS could not be coded had to be excluded. Furthermore, the model tested the effect of VOICE, which could not be investigated in 5.2.2 and 5.2.3 due to its correlation with CLAUSE_TYPE.

At the same time, FILLER_TYPE had to be excluded once again. *Which* clearly being the dominant filler type in finite RCs (cf. figure 5.1), there were only very few observations for other filler types. In addition, the almost categorical stranding preference

²⁸ The results obtained from this model are published in Mußemann (2024). However, as they are interpreted in a much broader context here, they are also included in this study.

of *who* and the strong pied-piping preference of *whom* would have led to many low-frequency cells (cf. table A.12 in the Appendix). Besides this, FILLER_TYPE is correlated with PP_TYPE as *which* typically appears in adjunct PPs, whereas *who*, *whom* and *where* are typical fillers of constructions with lexically entrenched V + P chunks (cf. figure A.13 in the Appendix). Excluding FILLER_TYPE was thus important to reduce multicollinearity.

Table 5.29: Stranded and pied-piped finite RCs across varieties (absolute and relative frequencies)

VARIETY (VARIETY_TYPE)	Stranding		Pied-piping		Total	
CanE (V)	37	10.88%	303	89.12%	340	100%
BrE (V)	69	9.11%	688	90.89%	757	100%
IrE (V)	56	11.69%	423	88.31%	479	100%
NZE (V)	81	18.45%	358	81.55%	439	100%
JamE (IV)	87	13.43%	561	86.57%	648	100%
SgE (IV)	67	21.47%	245	78.53%	312	100%
KenE (III)	49	11.34%	383	88.66%	432	100%
TanE (III)	31	8.64%	328	91.36%	359	100%
HKE (III)	35	12.46%	246	87.54%	281	100%
IndE (III)	31	4.86%	607	95.14%	638	100%
NigE (III)	63	16.15%	327	83.85%	390	100%
PhilE (III)	54	14.48%	319	85.52%	373	100%
Column total	660	12.11%	4,788	87.89%	5,448	100%

The variable FREQUENCY_FILLER also had to be removed from the model. As the filler types *how*, *wh-ever* and *when* only occur in free RCs and interrogatives, there are only five levels of FILLER_TYPE (*which*, *whom*, *who*, *whose*, *where*) that appear in the finite RC data of all varieties. Furthermore, as shown in table 5.30, *which* is by far the most frequent FILLER_TYPE in finite RCs. As most of the range of FREQUENCY_FILLER is thus hardly attested at all, this leads to a highly skewed distribution which cannot be improved by a log-transformation (cf. figure A.14 in the Appendix). Consequently, testing the effect of FREQUENCY_FILLER would not be very helpful. For example, if the model estimates suggested that frequent fillers prefer pied-piping, this would only reveal that *which*, i.e. the most frequent filler type, prefers pied-piping since the other filler types have much lower frequencies. Note that this was not a problem in the models that also included the free RC and interrogative data because in these clause types fillers such as *what* or *where* occur much more frequently so that the

overall distribution of `FREQUENCY_FILLER` is not as skewed (cf. table 5.30 and figure 5.17). In contrast, the variable had to be excluded from the regression analysis that only investigated finite RCs.

Table 5.30: Distribution of `FILLER_TYPE` in the finite RC data compared to the finite RC, interrogative, and free RC data (absolute and relative frequencies)

Finite RC data			Finite RC, interrogative, free RC data		
<code>FILLER_TYPE</code>	Abs. freq.	Rel. freq.	<code>FILLER_TYPE</code>	Abs. freq.	Rel. freq.
which	4,894	89.83%	which	5,077	56.57%
whom	318	5.84%	what	2,344	26.12%
where	101	1.85%	where	616	6.86%
who	79	1.45%	whom	371	4.13%
whose	48	0.88%	who	245	2.73%
what	5	0.09%	how	131	1.46%
whence	3	0.06%	wh-ever	98	1.09%
			whose	67	0.75%
			when	19	0.21%
			whence	3	0.03%
			whereabouts	2	0.02%
			why	1	0.01%

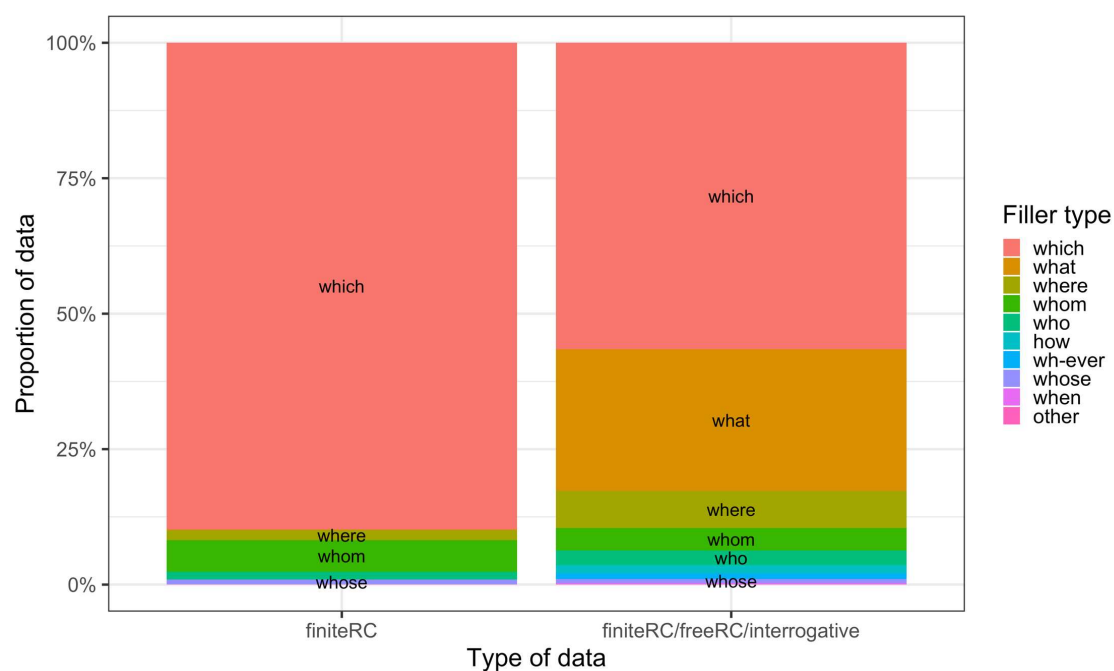


Figure 5.17: Distribution of `FILLER_TYPE` in the finite RC data compared to the finite RC, interrogative, and free RC data

Another factor that could not be investigated is FILLER_LENGTH. Since complex filler phrases typically occur in interrogative constructions (e.g. *Which school do you go to?*), the finite RC data contain very few tokens with fillers consisting of more than one word (cf. table 5.31). FILLER_LENGTH was thus not included in the regression analysis to avoid overfitting.

Table 5.31: Tokens with FILLER_LENGTH > 1 across variety phases (absolute and relative frequencies)

VARIETY_TYPE	Abs. freq.	Rel. freq.
III	27	54%
IV	7	14%
V	16	32%
Total	50	100%

Table 5.32: Recoding of PP_TYPE (finite RCs)

Factor level	PP type (token numbers)
lexicalized	VXP (73)
	prepositional X (682)
	obligatory complement with <i>be</i> (24)
complement_like	obligatory complement without <i>be</i> (57)
	optional complement (1,112)
	movement (168)
	accompaniment (65)
	subcategorized PP (85)
adjunct_like	affected location (66)
	direction (21)
	location (395)
	time (268)
	duration/frequency (136)
	means/instrument (364)
	agent (2)
	cause/reason/purpose/result (149)
	respect (1,139)
	manner (484)
	degree (158)

The other numeric predictors, i.e. FREQUENCY_PREPOSITION and COMPLEXITY, were log-transformed, centered and standardized. In addition, once again, PP_TYPE had to be recoded (cf. table 5.32). The lexicalized level and the complement_like level correspond to the coding used in 5.2.2 and 5.2.3. However, in contrast to the previous models, no distinction was made between the adjunct_like group and the respect_manner_degree group. Instead, subcategorized PPs and all adjunct PPs apart from movement and accompaniment PPs (which belong to the complement_like level) were assigned to the adjunct_like level to avoid factor combinations for which the data contain no observations. Since the present model only analyzes finite RCs, i.e. the most complex clause type, in which pied-piping should be the prototypical choice, following Jach (2021: 359), it is assumed that all adjunct PPs strongly favor pied-piping.

Finally, compared to the previous models, it was possible to implement a more nuanced formality classification since the maximal model did not have to include a complex three-way interaction between FORMALITY, CLAUSE_TYPE and VARIETY_TYPE. Recall that, according to Bohmann (2019: 194), varieties at lower Dynamic Model stages “tend towards linguistic patterns that are more formal and informational, and less affective and involved than their phase 5/Inner Circle counterparts” (cf. 2.3). Since such differences in style cannot be fully captured by a simple spoken-written or formal-informal dichotomy (Bohmann 2019: 195), the variable FORMALITY (cf. 5.2.2) was recoded into the factor TOPIC, which consists of a three-level distinction between “personal_less_formal”, “other_less_formal” and “more_formal” (also see Mußemann 2024).²⁹ This means that all texts categorized as less formal in section 5.2.2.1 were 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 Dynamic Model stages (Van Rooy et al. 2010: 346; cf. 2.3). As shown in table 5.33, unscripted monologues, public dialogues, private dialogues and private correspondence texts dealing with personal topics make up the personal_less_formal level. The other_less_formal level includes private dialogues about non-personal topics as well as private correspondence texts focusing on topics categorized as medium in terms of formality (i.e. e.g. social letters about education, society or language; cf. table 5.12). The more_formal level contains all texts which were classified as more formal in section 5.2.2.1. A more fine-grained distinction between

²⁹ Note that this three-level distinction could not be tested in the logistic regression analysis in 5.2.2 as the three-way interaction TOPIC * CLAUSE_TYPE * VARIETY_TYPE would have suffered from many low-frequency cells.

topics was not implemented to reduce multicollinearity. A simple two-way topic distinction (“personal” vs. “other”) was not considered sufficient because personal topics in finite RCs almost exclusively occur in less formal text types such as private dialogue. The three-level distinction thus ensures that a potential effect of the personal topic is not due to its correlation with informal text types. Table 5.34 shows that most finite RCs in the dataset come from more formal texts.

Table 5.33: Levels of TOPIC (Mußemann 2024: 14)

Personal_less_formal	Other_less_formal	More_formal
Texts about personal topics from the categories: Private correspondence Private dialogues Public dialogues Unscripted monologues	Private dialogues about non-personal topics Private correspondence about medium topics	Private correspondence about formal topics Public dialogues/unscripted monologues about non-personal topics Regardless of topic: - Scripted monologues - Broadcast news - Non-professional writing - Business correspondence - Academic writing - Non-academic writing - Reportage - Instructional writing - Persuasive writing - Creative writing - Only ICE-EA: Written as spoken, Legal presentations

Table 5.34: Distribution of TOPIC across VARIETY_TYPE (absolute and relative frequencies)

VARIETY_TYPE	personal_less_formal		other_less_formal		more_formal		Total	
III	49	1.98%	78	3.15%	2,346	94.86%	2,473	100%
IV	35	3.65%	41	4.27%	884	92.08%	960	100%
V	118	5.86%	25	1.24%	1,872	92.90%	2,015	100%
Column total	202	3.71%	144	2.64%	5,102	93.65%	5,448	100%

Since adjunct PPs tend to be longer than prepositional verbs, the maximal model included one three-way interaction, namely VARIETY_TYPE * PP_TYPE * COMPLEXITY. In addition, the model tested the two-way interactions between VARIETY_TYPE, and the variables VOICE, PHRASE_TYPE, RESTRICTIVENESS,

FREQUENCY_PREPOSITION and TOPIC. Ideally, the maximal model would have also included the three-way interaction VARIETY_TYPE * VOICE * TOPIC to capture the fact that the passive is typically associated with formal contexts. However, this would have led to overfitting due to the low frequencies of less formal passive tokens (cf. table A.15 in the Appendix). Other potentially interesting three-way interactions (such as VARIETY_TYPE * TOPIC * PP_TYPE) also had to be excluded due to data sparsity.

With respect to the random effects, the maximal model included random intercepts for FILE_ID and PREPOSITION (cf. 4.4.1). Prepositions with ≤ 5 observations ($N = 37$) were again combined into one group “other”. A more complex random effects structure was not considered because of convergence problems.

The final model structure, which only includes random and fixed effects that significantly improve the model fit, is given in (5.30):³⁰

$$\begin{aligned}
 (5.30) \text{ PREPOSITION\_PLACEMENT} \sim & \\
 & \text{VARIETY\_TYPE} + \text{PP\_TYPE} + \text{COMPLEXITY} + \text{VOICE} + \\
 & \text{PHRASE\_TYPE} + \text{RESTRICTIVENESS} + \text{TOPIC} + \\
 & \text{PP\_TYPE:COMPLEXITY} + \\
 & \text{VARIETY\_TYPE:RESTRICTIVENESS} + \\
 & \text{VARIETY\_TYPE:TOPIC} + \\
 & (1 \mid \text{FILE\_ID}) + (1 \mid \text{PREP})
 \end{aligned}$$

This model is highly significantly better ($p < 0.001$) than a null model which does not include any fixed effects. It has a marginal R^2 of 0.51 and a conditional R^2 of 0.74. All adjusted GVIFs are below 2, and the condition number K is 11.36, which suggests that there is no harmful collinearity. Moreover, the classification accuracy of the final model is 95.06% (precision = 95.38%, recall = 99.19%), which is highly significantly better than a model that always predicts the most frequent choice ($p < 0.001$). The index of concordance $C = 0.98$, which also indicates that the model has excellent predictive capacities.

³⁰ Compared to the models in 5.2.2 and 5.2.3, this time, LRTs with a significance threshold of $p < 0.1$ were used to identify the final model structure since this is, to my knowledge, the first study to explore the effect of TOPIC on preposition placement in different variety phases. A less conservative significance threshold of $p < 0.1$ was thus used to decrease the Type II error rate (see Gries & Wulff 2013: 338; Gries 2021: 773).

5.2.4.2 Results

The random intercept estimates for FILE_ID and PREPOSITION are listed in table 5.35. Figure 5.18 visualizes the intercept adjustments for the various prepositions, with positive estimates indicating a stranding preference and negative estimates indicating a pied-piping preference.

Table 5.35: Variance estimates and standard deviations of random effects (of the final model for the finite RC data)

Groups	Variance	SD
FILE_ID	1.878	1.37
PREPOSITION	1.01	1.005

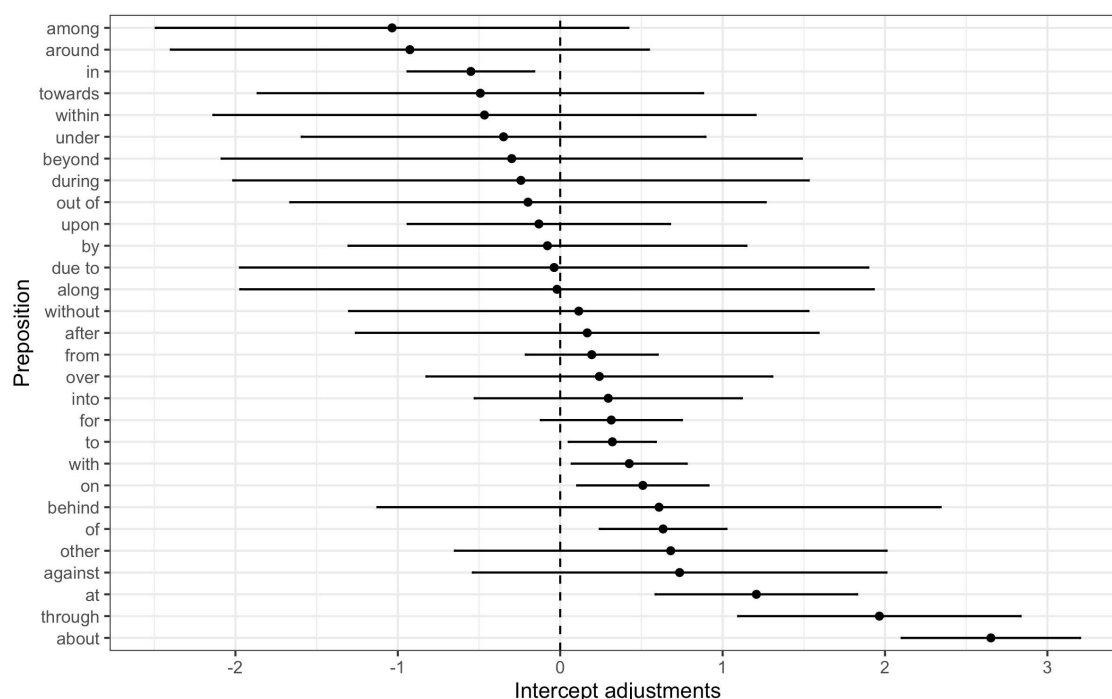


Figure 5.18: Intercept adjustments of PREPOSITION and Wald confidence intervals (of the final model for the finite RC data)

That *about* is again the preposition with the strongest stranding probability provides further support for the entrenchment of a stranded *WH...about* meso-construction, which is likely to emerge as a result of the frequent occurrence of stranded constructions with optional complement PPs that express their theme with an *about*-phrase (e.g. *talk/complain/trumpet about* in (5.31)). This effect will also be explored in sections 5.3.1 and 5.3.2. The strong stranding preference of *at* can be explained by the prepositional verb *look at* ‘examine, study’, which makes up 25 of all 39 stranded finite RC constructions with *at* in the dataset (5.32).

- (5.31) (a) *Right now I have published thirteen books twelve plays and one novel plus a novella which is in Kimaragoli **which I talked about*** <ICE-EA-K:S1BINT7K>
- (b) *And if the stories are true then I have ingested spit from every single waiter in Metro Manila **whose service I've complained about*** <ICE-PHI:S1B-035#11:1:C>
- (c) *and the criminal justice act **which he trumpeted about*** <ICE-NZ:S1B-056#74:1:P>
- (5.32) (a) *But <,> on this particular issue <,> there are four categories of public officers **whom I think we could look at*** <ICE-JA:S1B-022#59:1:C>
- (b) *um so the first first two questions are just basic age gender stuff um **which i'll look at** to see if that has any effect on my results* <ICE-NZ:S2A-040#33:1:A>

With respect to the prepositions which favor pied-piping, the results are very similar to those of the regression models investigating all clause types. That *since*, which turned out to have a very strong pied-piping preference in 5.2.2 and 5.2.3, is not included in figure 5.18 can be explained by the fact that *since* typically occurs in interrogative constructions.

The results for the fixed effects are given in tables 5.36 and 5.37, in which a positive *b* value indicates that a level change of a categorical predictor makes stranding more likely, while a negative *b* value reflects a decreasing probability of preposition stranding. As the interactions of VOICE, PHRASE_TYPE, PP_TYPE and COMPLEXITY with VARIETY_TYPE did not emerge as significant, these four processing-related variables can be said to affect all variety phases in a similar way. With respect to VOICE, the model indicates that preposition stranding is negatively associated with the passive, which confirms Gries' (2002) finding that preposition stranding is more likely when the main verb of a construction is in the active. The semantic roles in passive constructions being expressed in a different syntactic order than in the more frequent active, the passive can be said to involve more processing cost (cf. 3.2.5). That passives favor pied-piping, i.e. the variant that reduces processing load, may thus reflect a processing principle. Besides this, the model shows that PPs contained in NPs are less likely to be stranded than VP- or AdjP-contained PPs, while there is no significant difference between PPs embedded in VPs and AdjPs. Once again, the logistic regression analysis thus indicates that speakers of all variety phases avoid stranded constructions with NP-contained PPs, in which speakers cannot establish a structural relation between the filler and the gap after encountering the main verb (cf. 5.2.2 and 5.2.3).

Table 5.36: Coefficients of significant main effects (of the final model for the finite RC data). Model predictions are for preposition stranding.

Predictors	<i>b</i>	SE	<i>z</i>	<i>p</i>
(Intercept)	-1.0192	0.3056	-3.335	<0.001
VARIETY_TYPE				
V > III	0.2931	0.1914	1.531	0.1258
V > IV	0.6775	0.2319	2.9216	<0.01
PP_TYPE				
lexicalized > complement_like	-0.9011	0.1956	-4.606	<0.001
lexicalized > adjunct_like	-4.3859	0.3069	-14.2903	<0.001
COMPLEXITY (logged)				
one-unit increase	-0.128	0.1539	-0.8313	0.4058
VOICE				
active > passive	-2.1496	0.2564	-8.3848	<0.001
PHRASE_TYPE				
VP > AdjP	-0.3963	0.254	-1.5602	0.1187
VP > NP	-1.5498	0.3064	-5.0588	<0.001
RESTRICTIVENESS				
restrictive > non-restrictive	0.7232	0.2629	2.7511	<0.01
TOPIC				
more_formal > other_less_formal	2.2438	0.7807	2.874	<0.01
more_formal > personal_less_formal	3.3642	0.418	8.049	<0.001

Table 5.37: Coefficients of significant interaction effects (of the final model for the finite RC data). Model predictions are for preposition stranding.

Predictors	<i>b</i>	SE	<i>z</i>	<i>p</i>
(Intercept)	-1.0192	0.3056	-3.335	<0.001
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	<0.05
VARIETY_TYPE * RESTRICTIVENESS				
III + non-restrictive	-0.9285	0.3651	-2.543	<0.05
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	<0.01
IV + personal_less_formal	-0.3074	0.7376	-0.4168	0.6768

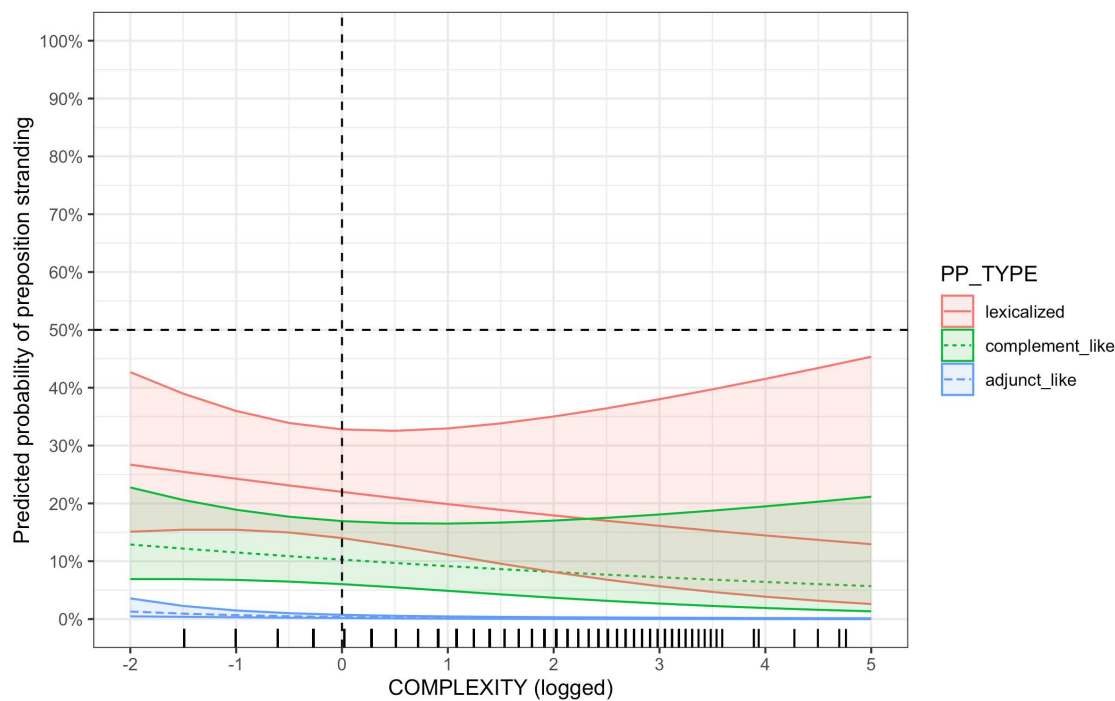


Figure 5.19: The effect of *COMPLEXITY (logged)* * *PP_TYPE*

All other predictors are involved in interactions (cf. table 5.37). While complement_like and lexicalized PPs generally favor stranding more than adjunct_like PPs, the interaction of PP_TYPE and COMPLEXITY indicates that not all PP types are affected equally by increasing complexity. As shown in table 5.37, a one-unit increase in complexity (logged) makes preposition stranding significantly less likely for adjunct_like PPs than for lexicalized PPs. This effect is also represented in figure 5.19: pied-piping is generally preferred in all finite 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 relative clauses.

Note that the finding that V + P chunks favor stranding less with increasing complexity is contrary to the predictions of Trotta (2000: 188) and Hoffmann (2011: 168; 224–225), who predicted that stranding should be strongly preferred with entrenched verb-preposition strings in longer FGDs. Although Hoffmann and Trotta correctly point out that in very long constructions breaking up lexicalized verb-preposition combinations could lead to an increase in processing efforts, in line with Hawkins' (1999: 251) “Minimize FGDs” principle, which claims that “[t]he human processor prefers FGDs to be as small as possible”, the great complexity resulting from long FGDs and PP-contained gaps might lead to an avoidance of stranding in such highly complex environments.

Complex RCs with e.g. prepositional verbs being an infrequent phenomenon (also see the large confidence band in figure 5.19), 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 relative constructions as this study. Support for this claim comes from the boxplots in figure 5.20, which show that the data for BrE and KenE contain shorter relative clause constructions than e.g. the data for TanE, IndE or PhilE. Not only is the median in the latter ICE components greater, but also there are more long outliers. Moreover, compared to Hoffmann’s data, the complex relative clauses 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.

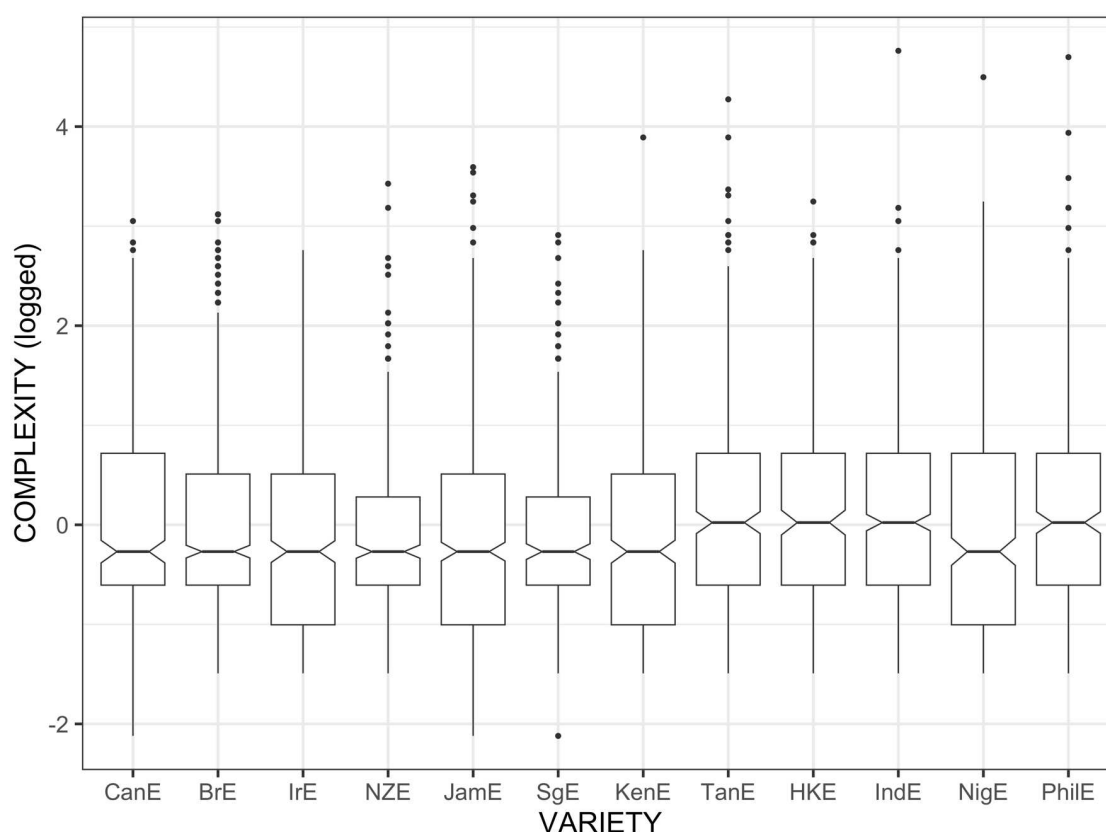


Figure 5.20: Boxplot: *COMPLEXITY* (logged) across *VARIETY*

Furthermore, figure 5.19 shows that all types of adjunct PPs (apart from movement and accompaniment adjuncts) show an almost categorical pied-piping probability in finite RCs, which can be attributed to the high processing complexity of this construction type. Interestingly, not even the PP types with the strongest associations between verb and prepositions, i.e. the groups of lexicalized and complement_like PPs,

favor stranding. That even these PP types generally prefer pied-piping may reflect a formality effect because speakers might consciously avoid stranding with e.g. 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 the logistic regression analysis.

Now, these results, of course, raise the question why the interaction of PP_TYPE and VARIETY_TYPE did not emerge as significant in the present analysis even though variety-specific effects of PP_TYPE were identified in the regression models in 5.2.2 and 5.2.3. As pointed out above, the stronger pied-piping preference with adjunct_like PPs in the less advanced varieties found in 5.2.2 and 5.2.3 may indicate that varieties at lower Dynamic Model stages have a stronger preference for constructions that are easy to process. One reason why no cross-varietal differences regarding PP type were identified in the present analysis could thus be that this regression model focused on preposition placement in finite 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 e.g. 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” (cf. 2.4.1.1). Importantly, though, this explanation cannot account for the fact that Hoffmann’s (2011) study found that KenE speakers rate stranding in finite RCs with prototypical adjunct PPs lower than BrE speakers. One reason for the differences between this study and Hoffmann’s analysis 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 speakers of more advanced varieties. 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).

Another reason why the present model did not identify a significant interaction effect of PP_TYPE and VARIETY_TYPE could be that pied-piping is the prototypical choice in finite RCs. Consequently, 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.

Finally, note that the fact that the interaction of `VARIETY_TYPE` and `PP_TYPE` did not emerge as significant in the present regression model provides support for the suggestion made in 5.2.3 that the stronger pied-piping preference with adjunct_like PPs in the less advanced varieties might be particularly relevant in connection with interrogatives. This claim about a potential interaction effect of `PP_TYPE`, `CLAUSE_TYPE` and `VARIETY_TYPE` will be investigated further with the help of an HCFA in section 5.3.2.

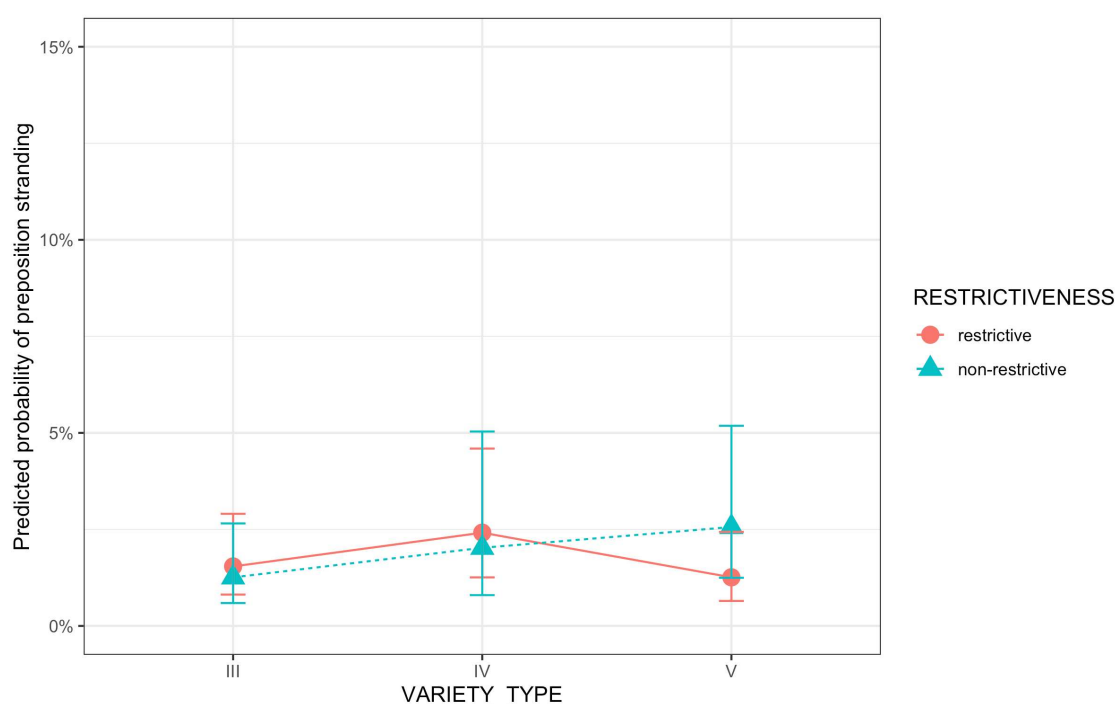


Figure 5.21: The effect of *RESTRICTIVENESS* * *VARIETY_TYPE*

While the logistic regression analysis did not identify a significant interaction of `PP_TYPE` and `VARIETY_TYPE`, two other interactions involving `VARIETY_TYPE` turned out to be significant. 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 figure 5.21, in both restrictive and non-restrictive relative clauses, Phase III exhibits a pied-piping probability that is comparable

to that of restrictive relative clauses in Phase V, while in Phase IV both restrictive and non-restrictive relative clauses show a pied-piping preference that is at a similar level as that of non-restrictive relative clauses in Phase V.

This interaction of `VARIETY_TYPE` and `RESTRICTIVENESS` provides at least some evidence for the claim that processing constraints have the strongest effect on varieties at lower stages of evolution. Recall that restrictive relative clauses, in which the interpretation of the antecedent noun is dependent on the parsing of the relative clause, are harder to process than non-restrictive relative clauses (cf. 3.2.1). Now, in the least complex clause type investigated, i.e. in non-restrictive relative clauses, Stage III displays the strongest preference for the prototypical pied-piping construction. This effect can be interpreted as an indication that the constructional choices of less advanced varieties are more strongly guided by processing factors than those of more advanced varieties. At the same time, surprisingly, Phase IV uses even more stranded prepositions than Phase V in restrictive relatives, which cannot be explained from a processing perspective. Even though the variable `VARIETY` could not be included in the present model, taking into account the results of section 5.2.3.2, this weaker pied-piping preference of Phase IV may be attributed to a stronger preference for stranded finite RC constructions in SgE.

Regarding Phase III, the interaction of `RESTRICTIVENESS` and `VARIETY_TYPE` thus supports the claim that speakers of varieties at lower Dynamic Model stages have the strongest preference for constructions that are easy to process. It must be kept in mind, though, that this effect is rather weak (cf. figure 5.21) and that no other processing-related variables turned out to have a stronger effect on varieties at earlier stages of evolution. Disregarding the higher preference for stranding in non-restrictive relative clauses in Stage III, with respect to finite RCs, the logistic regression analysis thus suggests that processing factors drive the constructional choices of speakers in very similar ways in varieties of English around the world.

While processing constraints are highly relevant in all varieties, the significant interaction of `VARIETY_TYPE` and `TOPIC` indicates that there are clear differences between variety stages with respect to topic, which has a much stronger effect on variety types IV and V. The significant negative coefficient of the interaction in table 5.37 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 5.22, both Stages IV and V display an almost categorical pied-piping preference in more formal contexts, but exhibit an 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 formal text types than in more formal ones, and the distinction between personal and non-personal topics does not make a difference at all.

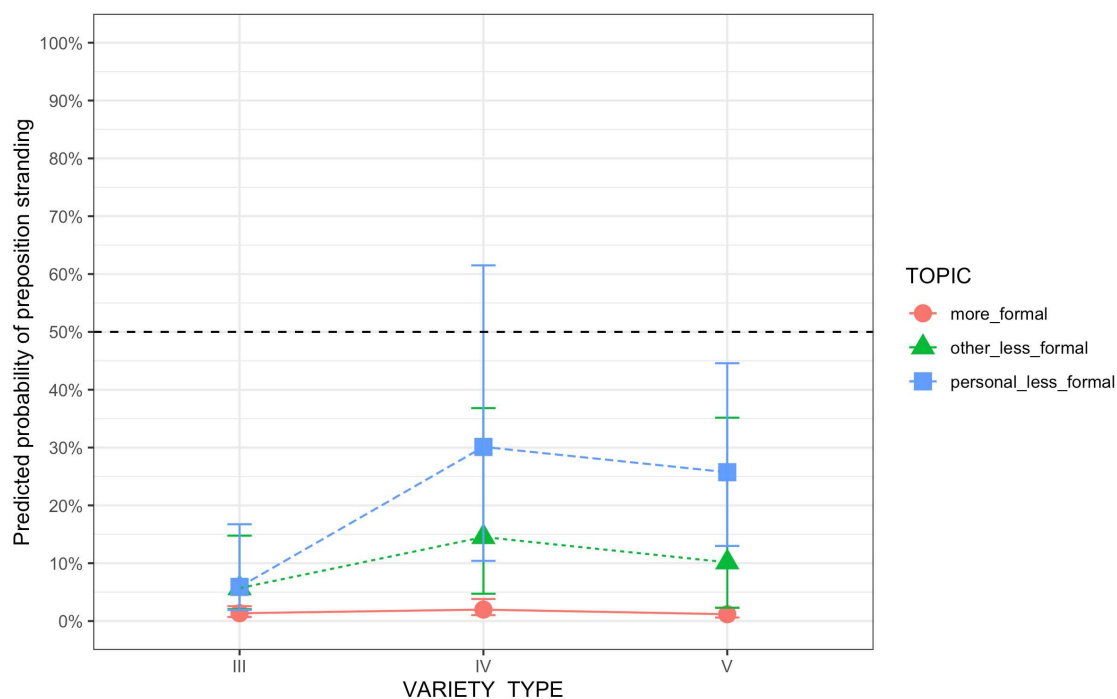


Figure 5.22: The effect of *VARIETY_TYPE* * *TOPIC*

The study thus corroborates Hoffmann's (2011) results, who found a strong formality effect for BrE relative clauses but only a weak one for KenE. Furthermore, the results show 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. Now, recall that the regression model in 5.2.2.2 did not find clear differences between variety phases regarding preposition placement in finite RCs, while the analysis in 5.2.3.2 identified cross-varietal differences that could neither be explained by processing factors nor by cross-linguistic influence. The present analysis indicates that these effects might actually be skewed by an uneven distribution of topics and formality levels across the various ICE components which could not be accounted for in the previous sections.

As pointed out in 3.3.1, Hoffmann (2011: 267–268) 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. In line with the claims made in 2.3, an input-based explanation can

also account for the lack of register distinctions 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 relative clause 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, 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 relative clauses 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-relative clause 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, as shown in section 3.1, 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. Instead, figure 5.22 indicates that Phase-III speakers only possess an abstract pied-piped finite WH-relative clause construction, which is used regardless of formality and topic. Furthermore, considering the frequent use of pied-piped adjunct PPs in earlier Dynamic Model stages (cf. 5.2.2.2 and 5.2.3.2), speakers of less advanced varieties can generally be expected to encounter a lot of positive input for pied-piped prepositions. Hence, they can be assumed to store many low-level pied-piping constructions which have a high chance of being activated across all clause types and registers. The combination of both factors, i.e. the pied-piping preference with adjunct PPs, which can be linked to processing factors, and the formal input which is typical of Stage-III varieties, can thus explain why the logistic regression analysis did not identify a strong formality effect for Phase III.

That Phase IV is more strongly affected by formality and topic than Phase III is consistent with studies such as Gut & Coronel (2012) or Hundt (2006), who found a greater degree of stylistic variation in Stage-IV varieties such as SgE and JamE than in Stage-III varieties such as PhlE or NigE. At the same time, it is slightly surprising that Phase IV turned out to pattern with Phase V. Nevertheless, taking into account the cognitive explanation of stylistic variation in PCEs provided in 2.3, this effect can also be linked to input. Speakers of Stage-IV varieties also frequently encountered the pied-piped construction in the formal domains of use through which the English language was

transplanted to the British colonies. As a consequence, associations between the pied-piped relative clause construction and formal contexts became entrenched in the minds of these speakers. However, since Stage IV tends to use even slightly more stranded relative clause constructions than Stage V, it is plausible to conclude that speakers of Stage-IV varieties also have an entrenched abstract stranded finite WH-relative construction. As a result of statistical preemption, this stranded construction became associated with more informal situations (for a similar explanation regarding BrE, see Hoffmann 2011: 269–269). Drawing on terminology from Schmid’s (2020) Entrenchment-and-Conventionalization Model (cf. 2.3), this means that pragmatic associates that link the stranded relative clause construction to informal contexts and situations in which speakers recount personal experiences became routinized and entrenched in the minds of individual speakers. 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. 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.

5.2.4.3 Summary

In contrast to the models investigating all clause types (cf. 5.2.2 and 5.2.3), the logistic regression analysis which only included the relative clause data did not identify clear processing-related differences between variety phases. This effect can be attributed to the high processing complexity of finite RCs, which may lead to similar constructional choices across all variety types. With respect to hypothesis I-III, the model yielded the following results:

Hypothesis I was confirmed because all variety phases prefer constructional choices which are preferred from a processing perspective, i.e. pied-piping with long FGDs, with NP-embedded PPs, with adjunct PPs, and with constructions in which the main verb is used in the passive.

Hypothesis II was not confirmed because, apart from a minor interaction effect of VARIETY_TYPE and RESTRICTIVENESS, the regression analysis did not

identify any significant interactions of VARIETY_TYPE and processing-related factors.

Hypothesis III was confirmed. Speakers of more advanced varieties exhibit an almost categorical pied-piping preference in more formal contexts, but they display an increasing stranding probability in less formal texts about non-personal topics, and the strongest stranding preference in less formal contexts dealing with non-personal topics. At the same time, the pied-piping probability of Phase III is only slightly weaker in less formal texts than in more formal ones. However, the results did not confirm a direct correlation between the strength of formality effects and Dynamic Model stages since Stage IV turned out to pattern with Stage V.

This section used generalized linear mixed-effects modeling to explore how processing constraints, stylistic variables and cross-linguistic influence affect the choice between stranding and pied-piping in twelve varieties of English. While the study has so far focused on an abstract constructional level, in the next section, HCFAs will be performed to identify meso-constructions that are significantly associated with individual varieties and variety phases.

5.3 Constructional fillers across varieties

In order to test the claim that varieties at earlier Dynamic Model stages rely on a small set of frequent and prototypical meso-constructions (cf. hypothesis IV), a number of HCFAAs were conducted which tested associations between fillers of constructional slots and specific varieties or variety phases. This method was used to identify entrenched V + P patterns (cf. 5.3.1), to analyze the use of P + FILLER and FILLER + P combinations in interrogative constructions (cf. 5.3.2), and to identify entrenched (ANTECEDENT +) P + WH strings in finite RCs (cf. 5.3.3). In addition, a closer look was taken at all pied-piped free RCs (cf. 5.3.4).

5.3.1 V + P combinations

From a corpus-as-input point of view, the results of the logistic regression analysis suggest that speakers of varieties at lower stages of evolution receive less input for stranded prepositions in free RCs and interrogatives than speakers of more advanced varieties, while there are no clear differences between variety phases regarding finite RCs. In order to assess whether this effect is also reflected in the entrenchment of V + P³¹ combinations, at first, an HCFA was conducted which tested the associations between the V slot, the P slot and the variables VARIETY/VARIETY_TYPE in free RCs and interrogatives. As some of the configurations are based on very low frequencies, in the following, only significant types with a *Q* value of at least 0.005 will be included. All significant configurations are given in tables A.16 and A.17 in the Appendix.

As shown in table 5.38, the HCFA identified slightly more entrenched V + P combinations for Phase V (9) than for Phase III (8), which might be taken as an indication that Phase V receives more input for stranded constructions. Note that the lower number of significant types involving Phase IV (5) can be explained by the fact that Stage IV only includes two varieties (JamE and SgE), which means that the overall token numbers of Stage IV are lower (cf. table A.18 in the appendix). While there are thus no major quantitative differences between variety phases, a few qualitative aspects deserve further comments. First of all, the V + P combinations associated with Phase V include two verb-preposition strings that consist of *be*, namely *be like* and *be up to*. Both the *WH...be like* and the *WH...be up to* meso-construction have non-compositional meanings. While

³¹ The term “V + P” does not only refer to verb-preposition strings but also to adjective-preposition combinations (e.g. *capable of*, *familiar with*). However, for the sake of simplicity, the expression V + P will be used throughout this study.

WH...be like expresses the meaning ‘how’ (cf. 3.1.2), *WH...be up to* (5.33) can be paraphrased by *What is X doing?*.

- (5.33) (a) *Well, **what** have you been **up to**?* <ICE-GB:W1B-005 #111:5>
 (b) ***what** have you been **up to** over the holidays and stuff* <ICE-NZ:S1A-093#1:1:S>

Table 5.38: HCFA results for *VARIETY_TYPE* * *V* * *P* in free RCs/interrogatives (only significant types with $Q \geq 0.005$ are shown)

VARIETY_TYPE	V	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
V	be	like	184	68.7411	193.2558	>	5.60834256501896E-27	***	0.032
V	talk	about	132	37.0024	243.8908	>	9.93432229684707E-30	***	0.026
V	come	from	101	13.2835	579.2287	>	6.6779643097573E-49	***	0.024
V	look	like	72	19.9336	135.997	>	4.09867891727612E-15	***	0.014
V	look	at	27	6.6076	62.9351	>	6.83679795675268E-05	***	0.006
V	get	from	28	5.1214	102.2045	>	5.212105470949E-08	***	0.006
V	think	of	21	0.7898	517.159	>	1.98686891794585E-18	***	0.006
V	go	to	27	6.5783	63.3972	>	6.23473640501451E-05	***	0.006
V	be	up to	30	6.7468	80.1434	>	1.26588100510945E-06	***	0.006
IV	be	like	94	32.3708	117.3329	>	2.099187E-14	***	0.017
IV	talk	about	73	17.4247	177.2549	>	6.76280702253076E-19	***	0.015
IV	come	from	46	6.2553	252.5284	>	4.45823922064727E-20	***	0.011
IV	look	at	21	3.1116	102.8393	>	7.068738949229E-07	***	0.005
IV	go	to	20	3.0978	92.2217	>	4.47564934061265E-06	***	0.005
III	talk	about	156	27.8303	590.2729	>	9.83981820316621E-60	***	0.035
III	come	from	82	9.9908	519.01	>	1.62149462879131E-41	***	0.02
III	look	like	57	14.9925	117.7009	>	2.7554034E-12	***	0.011
III	refer	to	25	1.5607	352.022	>	2.94376980964827E-17	***	0.006
III	look	at	25	4.9697	80.7318	>	4.32913648781489E-06	***	0.005
III	ask	for	20	1.8927	173.231	>	7.3260986744E-10	***	0.005
III	stand	for	21	1.7508	211.6357	>	1.455407959E-11	***	0.005
III	go	to	25	4.9477	81.269	>	3.95570709202015E-06	***	0.005

Since these meso-constructions occur more frequently than expected by chance only in the Phase-V varieties (and *be like* in Phase IV), it can be said that these idiomatic constructions are less deeply entrenched in varieties at lower developmental stages. At the same time, the only entrenched partly schematic construction involving *like* in Phase III is the chunk *look like*. This preference for *look like* over *be like* in the less advanced varieties will be addressed again in section 5.3.2.

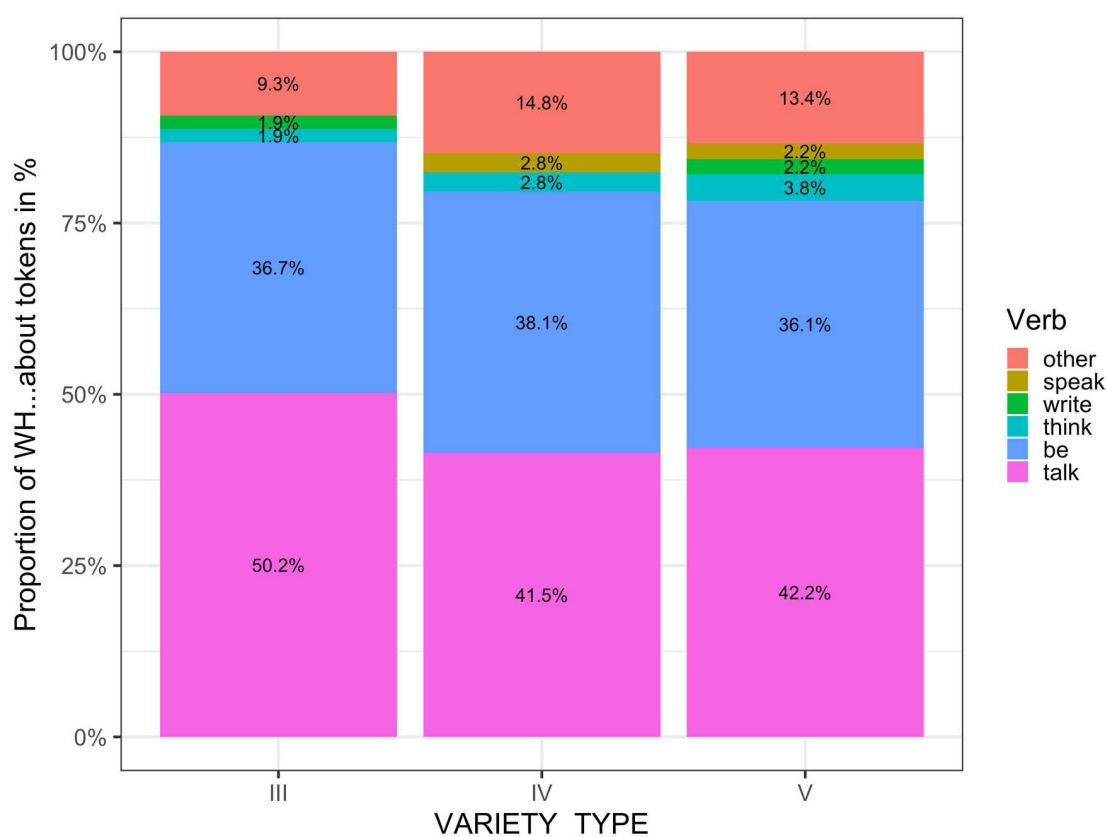


Figure 5.23: Fillers of the *V* slot in *WH...about* free RC/interrogative constructions across variety phases

While *be like* is the *V + P* chunk with the highest *Q* value in Stages IV and V, the *V + P* combination with the largest effect size in Phase III is *talk about*, which makes up 12.57% of all Stage-III tokens included in the HCFA (vs. 8% in Stage V, and 9.4% in Stage IV). The fact that *talk about* turned out to be significantly associated with all variety phases provides support for the claim made in 5.2.4.2 that the strong stranding preference of *about* can be attributed to the frequent occurrence of *about* in constructions with optional complement PPs. At the same time, a closer look at the verbs used in the *WH...about* tokens reveals that there is more variation in the verb slot in the more advanced varieties. As shown in figure 5.23, more than half of the *WH...about* tokens in Phase III consist of the prototypical verb *talk*. Although the frequencies of the *WH...about* tokens are too low to be investigated with the help of an HCFA (as this statistical

technique is very conservative with small sample sizes; cf. 4.4.2), this indicates that speakers of varieties at higher stages of evolution may have entrenched a more abstract *WH...about* meso-construction, whereas speakers of Phase-III varieties may rely more on the more specific *WH...talk about* construction in which the V slot is fully specified. Likewise, the fact that both *come from* and *get from* are entrenched in Phase V, while Stages III and IV only have the prototypical chunk *come from* may be taken as an indication that Phase V exhibits more variation in the verb slot of a partly schematic *WH...from* construction.

Table 5.39: HCFA results for VARIETY * V * P in free RCs/interrogatives (only significant types with $Q \geq 0.005$ are shown)

VARIETY	V	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
CanE	talk	about	38	8.903	95.0955	>	4.687397678001E-08	***	0.008
CanE	come	from	25	3.1961	148.7469	>	1.50482592533E-09	***	0.006
KenE	come	from	20	1.7711	187.6194	>	8.714242497E-10	***	0.005
BrE	be	like	50	15.748	74.4983	>	5.4953214918127E-07	***	0.009
BrE	talk	about	29	8.4769	49.6877	>	0.003309921217712	**	0.006
HKE	talk	about	30	4.7318	134.9343	>	8.2815970495E-10	***	0.007
IrE	be	like	51	16.9561	68.3522	>	2.25780265860217E-06	***	0.009
IrE	talk	about	30	9.1272	47.7336	>	0.004547250993103	**	0.006
IrE	come	from	22	3.2766	106.9907	>	1.05451248655935E-06	***	0.005
JamE	be	like	62	18.2893	104.4668	>	9.157963689E-11	***	0.012
JamE	come	from	37	3.5342	316.892	>	5.03409702566141E-20	***	0.009
JamE	talk	about	34	9.8449	59.2661	>	0.000174730048124	***	0.007
NigE	talk	about	40	6.7277	164.5504	>	2.5442748E-13	***	0.009
NigE	come	from	23	2.4152	175.4447	>	3.0158006079E-10	***	0.006
NZE	come	from	37	3.7677	293.1193	>	4.288091923734E-19	***	0.009
NZE	talk	about	35	10.4952	57.2152	>	0.000243491926377	***	0.007
PhilE	talk	about	50	7.1089	258.7807	>	1.22611948065102E-20	***	0.012
PhilE	come	from	20	2.5521	119.2858	>	6.2110519180421E-07	***	0.005
SgE	talk	about	39	7.5799	130.2422	>	7.077466386E-11	***	0.009

These results are also confirmed when looking at the V + P combinations which are significantly associated with individual varieties (cf. table 5.39). While *be like* is

entrenched in more advanced varieties, such as the Stage-V varieties BrE or IrE and the Stage-IV variety JamE, *talk about* occurs more frequently than expected by chance in all varieties except KenE, TanE and IndE. The absence of a significant association between *talk about* and KenE/TanE can be explained by the smaller private dialogue section of ICE-EA, which may contain fewer *WH...talk about* constructions than the other ICE components (cf. 4.1). The fact that *talk about* did not turn out to be entrenched in IndE may be due to the low token numbers of this variety (131; cf. table A.18 in the Appendix), which is consistent with the finding that IndE generally favors pied-piping over stranding (cf. 5.2.3.2).

In addition, out of all varieties, it is NigE (11.08%) and PhilE (13.85%) that have the largest proportion of free RC/interrogative tokens involving the string *talk about* (cf. table 5.40), which provides further support for the claim that less advanced varieties rely heavily on an entrenched *WH...talk about* meso-construction.

Table 5.40: WH...talk-about free RC/interrogative tokens per VARIETY (absolute and relative frequencies)

VARIETY (VARIETY_TYPE)	Abs. freq.	Rel. freq.
CanE (V)	38	10.53%
BrE (V)	29	8.03%
IrE (V)	30	8.31%
NZE (V)	35	9.70%
JamE (IV)	34	9.42%
SgE (IV)	39	10.80%
KenE (III)	20	5.54%
TanE (III)	5	1.39%
HKE (III)	30	8.31%
IndE (III)	11	3.05%
NigE (III)	40	11.08%
PhilE (III)	50	13.85%

Next, an HCFA was conducted which tested associations between the V slot, the P slot and the variables VARIETY/VARIETY_TYPE in the finite RC data. As the frequencies for stranded finite RCs are even lower (cf. table A.19 in the Appendix), in the present discussion, only significant types with a *Q* value of at least 0.01 will be included. All significant configurations are given in table A.20 in the appendix.

Table 5.41: HCFA results for *VARIETY_TYPE* * *V* * *P* in finite RCs (only significant types with $Q \geq 0.01$ are shown)

<i>VARIETY_TYPE</i>	<i>V</i>	<i>P</i>	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	<i>Q</i>
V	look	at	11	0.9547	105.6961	>	0.000105215569655	***	0.016
V	know	about	10	0.8813	94.35	>	0.000594345195716	***	0.015
III	come	from	17	0.9608	267.7518	>	8.52803428E-12	***	0.026
III	talk	about	16	1.937	102.1001	>	4.7282646971247E-06	***	0.023
III	refer	to	11	1.7493	48.9198	>	0.040401956463288	*	0.015
III	rely	on	9	0.3792	195.9868	>	5.43121328860106E-06	***	0.014
III	look	at	9	1.0209	62.3627	>	0.022887777974805	*	0.013

Table 5.41 shows that the HCFA identified more significant *V* + *P* combinations for Phase III (5) than for Phase V (2). The fact that no meso-construction emerged as being significantly associated with Phase IV can, once again, be attributed to the low token numbers of Stage IV. Now, following Hoffmann's (2021: 21) Dynamic Model Productivity hypothesis, in 5.1 it was argued that more advanced varieties should have a greater number of entrenched meso-constructions because the productivity of constructions should be higher in more advanced varieties, whereas varieties at lower stages of evolution should rely more on a small number of frequent and prototypical meso-constructions. Note, however, that the fact that more *V* + *P* combinations turned out to be associated with Phase III does not necessarily imply that the pattern *V* + *P* is less productive in the Stage-V varieties. Recall that the previous HCFA (cf. table 5.38) indicated that Stage-V speakers store slightly more substantive *V* + *P* chunks for free RCs and interrogatives than speakers of Phase-III varieties. Furthermore, the LNRE models in section 5.4.1 will show that the *V* + *P* slot is generally more productive in Stage V. The combined results of tables 5.41 and 5.38 can thus be taken as an indication that speakers of more advanced varieties tend to rely on entrenched *V* + *P* chunks primarily in free RCs and interrogatives, which can be explained by the fact that many of the entrenched verb-preposition strings are restricted to these clause types (e.g. *be like*) or that they are semantically associated with interrogatives (e.g. speakers use chunks such as *come from* or *go to* to ask where somebody comes from or which school somebody goes to).

Considering that HCFAs tend to be conservative, the fact that table 5.41 includes only two entrenched V + P combinations for Phase V may thus indicate that speakers of the most advanced varieties use V + P combinations in finite RCs so productively that the frequencies of individual V + P chunks in the Phase-V data are so low that they do not emerge as significant. In contrast, the higher number of V + P chunks significantly associated with Phase III may suggest that speakers of varieties at earlier Dynamic Model stages, who do not have an abstract stranded finite WH-relative clause construction (cf. 5.2.4.2), use stranded finite RCs primarily with lexically entrenched V + P chunks as the use of these entrenched meso-constructions involves a reduction in processing cost. Support for this claim also comes from the fact that, despite the conservativeness of the HCFA, the V + P combinations significantly associated with Phase III make up almost a quarter of all stranded finite RC tokens in these varieties, whereas the meso-constructions that can be said to be entrenched in Stage V only account for 8.64% of all stranded finite RC tokens in Phase V (cf. table 5.42). While speakers of all variety phases seem to use lexicalized V + P chunks in stranded free RCs and interrogatives (cf. table 5.39), in stranded finite RCs, i.e. in non-prototypical stranding contexts, varieties at earlier Dynamic Model stages seem to rely more on a small set of specific meso-constructions (such as *WH...talk about* or *WH...rely on*).

Table 5.42: Proportion of entrenched V + P combinations of the overall number of stranded finite RC tokens in Phases III and V

VARIETY_TYPE	% of entrenched V + P combinations of the overall number of stranded finite RC tokens
III	23.57%
V	8.64%

Coming to the associations between V + P combinations and individual varieties (cf. table 5.43), the HCFA identified only two significant types with a *Q* value of at least 0.01: While *come from* is significantly associated with NigE, *rely on* occurs more frequently than expected by chance in HKE. Note that the fact that no associations between V + P combinations and more advanced varieties turned out to be significant cannot be attributed to low token numbers since BrE, NZE, SgE and JamE all have higher frequencies of stranded finite RC constructions than NigE and HKE (cf. table A.19 in the Appendix).

Table 5.43: HCFA results for VARIETY * V * P in finite RCs (only significant types with $Q \geq 0.01$ are shown)

VARIETY	V	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
NigE	come	from	7	0.2256	203.4242	>	0.000338569350873	***	0.011
HKE	rely	on	7	0.0537	898.5304	>	1.707473925621E-08	***	0.011

The results for NigE and HKE may explain why these varieties turned out to have a relatively strong preference for stranded finite RC constructions in the logistic regression analysis in section 5.2.3.2. The prepositional verb *rely on*, which occurs seven times in a single ICE-HK file, accounts for 20.00% of all stranded finite RC tokens in the ICE-HK data. Moreover, the V + P combination *come from* makes up 12.07% of all stranded finite RC tokens in NigE. This indicates that the stronger preference for stranded finite RCs in NigE and HKE revealed by the logistic regression analysis is not due to the entrenchment of a more abstract stranded finite WH-relative clause construction in these varieties (which is in accordance with the claim made in 5.2.4.2 that Phase III only has a pied-piped finite WH-relative clause construction). Instead, this effect can be attributed to the fact that these varieties rely more on entrenched V + P chunks.

5.3.2 Fillers of the P + FILLER/FILLER + P slots in interrogatives

Based on the results of the regression models, in 5.2.3.2 and 5.2.4.2 it was suggested that varieties at lower stages of evolution might have a stronger preference for pied-piped adjunct interrogative constructions than more advanced varieties. In order to assess this claim, an HCFA was performed which tested the associations between VARIETY/VARIETY_TYPE, the P slot and the FILLER slot in pied-piped interrogatives.

While the HCFA identified only 2 entrenched combinations for Stage V, there are 6 P + FILLER combinations that are significantly associated with Phase III (cf. table 5.44). This quantitative effect is in line with the regression results which suggested that Stage-III varieties favor pied-piping in interrogative constructions more than varieties at higher developmental stages. As predicted, the six P + FILLER combinations significantly associated with Stage III include a wide range of adjunct PPs, namely degree (*to what extent* (5.34)), duration/frequency (*for how long* (5.35), *since when* (5.36)), time

(*at what time* (5.37)) and movement (*from where* (5.38)) PPs, which suggests that there is indeed an interaction effect of variety type, clause type and PP type. The only exception is *of what use*, which typically occurs as an obligatory complement of *be* (5.39).

(5.34) *To what extent* does multilingualism affect the ethnic languages <ICE-EA-T:W1A007T>

(5.35) *for how long* have you been there <ICE-NIG:con_22:a118>

(5.36) And *since when* have you been the Trust Officer of Equitable P C I Bank <ICE-PHI:S1B-061#55:1:E>

(5.37) At *at what time* did you warn the departments <ICE-HK:S1B-028#125:1:B>

(5.38) but *from where* the water is going to come <ICE-IND:S1A-062#113:1:A>

(5.39) Indeed, counsel, advice and justice are founded on freedom, but *of what use* are these when man is not free regarding his birth? <ICE-PHI:W2A-009#128:1>

With respect to the variety-specific results, the HCFA only identified entrenched P + FILLER combinations for Phase-III (IndE, HKE, TanE) and Phase-IV (JamE) varieties (cf. table 5.45). Again, the significant types include degree (*to what extent*), duration/frequency (*for how long*, *on how many occasions*), time (*at what time*) and movement (*from where*) PPs.

Although these findings are based on rather low token numbers and the potential interaction effect needs to be corroborated by future research, this indicates that speakers of less advanced varieties prefer pied-piped interrogative constructions with adjunct PPs, which are easier to parse than stranded adjunct interrogative constructions. Note that the fact that the HCFA revealed only one significant combination for Stage IV does not necessarily mean that this effect does not also apply – at least to a certain extent – to Stage-IV varieties. Considering the fact that JamE and SgE also emerged as being more likely to use pied-piped prepositions with adjunct PPs than the Phase-V varieties (cf. figure 5.15), the absence of more significant associations involving Phase IV can again be explained by the small sample size of Stage IV, which subsumes only two varieties (cf. table 5.28). In contrast, speakers of Stage-V varieties seem to favor stranded interrogative constructions for discourse-functional reasons since stranding ensures that the WH-word appears clause-initially. The two significant combinations the HCFA identified for Stage V include *on what basis* (5.40a), i.e. a respect PP, as well as *at what point*, which is also frequently used in a fairly abstract sense to establish a point of

reference. For instance, in (5.41a), *point* does not refer to a time or location but a threshold of income.

- (5.40) (a) *on what basis* do you assume two hundred <ICE-GB:S1B-020 #153:1:A>
 (b) ? *what basis* do you assume two hundred *on*
- (5.41) (a) *at what point* does it it cut out that so it's your responsibility entirely to meet these social services <ICE-NZ:S2A-047#65:1:M>
 (b) ? *what point* does it cut out *at*

The HCFA thus indicates that speakers of the most advanced varieties use pied-piped interrogative constructions only with PPs with which stranding would be uninterpretable (5.40b and 5.41b).

Next, an HCFA of the stranded interrogative data was performed in order to identify FILLER + P combinations that are significantly associated with a variety phase or variety. Due to the rather low token numbers (cf. table A.21 in the Appendix), once more, only types with a *Q* value of at least 0.005 will be reported here, while all significant configurations are listed in tables A.22 and A.23 in the Appendix. The HCFA results in table 5.46 indicate that there are no major quantitative differences between variety phases (that fewer FILLER + P combinations are associated with Phase IV can again be explained by the low token numbers of this variety stage), which can be attributed to the fact that interrogative constructions are prototypical stranding contexts. Consequently, speakers of all variety types can be expected to receive a lot of positive input for stranded interrogative constructions, which should lead to the entrenchment of partly schematic FILLER + P constructions in the construction networks of all variety phases. At the same time, there are some interesting qualitative aspects which should be commented on briefly. While the string *for how long* turned out to be significantly associated with Phases III and IV in the HCFA of the pied-piped interrogative data (cf. table 5.44), the HCFA of stranded interrogatives shows that Stage-V speakers store the stranded variant *how...for* (5.42). While the string *how long...for* surfaces 37 times in the Phase-V data (and, as shown in (5.42), in all four Phase-V varieties), there are only three such tokens in the ICE components of Phase IV, and the Phase-III data do not contain a single instance of *how long...for* (cf. table 5.47).

- (5.42) (a) *How long* did you do English *for* <ICE-GB:S1A-006 #1:1:A>
 (b) *So how long* <,> did you see him *for* <ICE-CAN:S1B-019#29:1:A>

Table 5.44: HCFA results for VARIETY_TYPE * P * FILLER in pied-piped interrogatives (non-significant associations are not shown)

VARIETY_TYPE	P	FILLER	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
V	on	what_basis	6	0.2128	157.3857	>	0.000711548800805	***	0.015
V	at	what_point	6	0.2549	129.4868	>	0.002029187018272	**	0.015
IV	for	how_long	8	0.7319	72.1755	>	0.006901220189698	**	0.019
III	to	what_extent	37	9.4183	80.7736	>	2.137703684878E-08	***	0.073
III	for	how_long	22	3.1044	115.0121	>	1.341630737567E-08	***	0.049
III	at	what_time	10	1.2845	59.1358	>	0.006613488738614	**	0.023
III	from	where	8	0.4336	132.0351	>	0.000135711244123	***	0.02
III	since	when	5	0.1922	120.265	>	0.01246260710472	*	0.012
III	of	what_use	4	0.0526	296.2351	>	0.002060110183653	**	0.01

Table 5.45: HCFA results for VARIETY * P * FILLER in pied-piped interrogatives (non-significant associations are not shown)

VARIETY	P	FILLER	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
JamE	for	how_long	7	0.4164	104.0917	>	0.007840159765708	**	0.017
IndE	at	what_time	8	0.4386	130.3574	>	0.0005922103981	***	0.02
IndE	from	where	5	0.1481	158.953	>	0.014047738549142	*	0.013
HKE	to	what_extent	12	1.7611	59.5282	>	0.008929524431886	**	0.027
HKE	on	how_many_occasions	5	0.0491	499.2141	>	6.13653761491813E-05	***	0.013
TanE	to	what_extent	11	1.072	91.9451	>	0.000493849083827	***	0.026

Table 5.46: HCFA results for VARIETY_TYPE * FILLER * P in interrogatives (only significant types with $Q \geq 0.005$ are shown)

VARIETY_TYPE	FILLER	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
V	where	from	117	24.19	356.085	>	8.70568235557792E-40	***	0.041
V	what	like	202	120.9225	54.3618	>	1.77386685163E-09	***	0.037
V	how	for	46	7.6011	193.9819	>	2.0219285604745E-18	***	0.017
V	who	to	27	5.6857	79.9021	>	6.516057022329E-08	***	0.009
V	who	with	15	2.6426	57.786	>	0.000102846469409	***	0.005
IV	where	from	74	12.0596	318.1377	>	7.2966287654311E-31	***	0.027
IV	which	to	18	2.6469	89.0542	>	3.8011410540909E-07	***	0.007
III	what	about	170	92.7512	64.3375	>	8.44309816E-11	***	0.035
III	where	from	97	17.7346	354.2794	>	6.68717256696077E-37	***	0.035
III	how	like	29	6.9195	70.4601	>	2.3002511732751E-07	***	0.01
III	what	of	35	15.3147	25.3032	>	0.007757138105966	**	0.009
III	where	to	27	11.3098	21.7672	>	0.036742530148279	*	0.007
III	who	with	16	1.9374	102.0733	>	2.2533288828886E-07	***	0.006

- (c) *how long* do you leave it in the fridge **for** <ICE-NZ:S1A-087#231:1:B>
- (d) *Uhm how long* has he been in love with her **for** <ICE-IRE:S1A-018\$D>

Note that these results are in line with Hoffmann (2011: 145), who found that the interrogative string *how long...for* is entrenched in BrE, whereas his KenE data only contained the pied-piped variant *for how long*. Stranding with adjunct PPs being a non-prototypical constructional choice, this finding provides further support for the claim that less advanced varieties are more strongly affected by processing constraints than varieties at higher developmental stages.

Table 5.47: *how...for* tokens per VARIETY_TYPE (absolute frequencies)

VARIETY_TYPE	Abs. freq.
III	0
IV	3
V	37

Furthermore, the HCFA shows that the meso-construction *what...like* is entrenched in Phase V, while Stage III stores the pattern *how...like*. Recall that the HCFA in 5.3.1 revealed that Stage V stores the strings *be like* and *look like*, whereas only *look like* occurs more frequently than expected by chance in the Phase-III varieties. Taken together, these results suggest that less advanced varieties have entrenched a *how...look like* construction, whereas Stage V has a more abstract *what...V like* construction in which the V slot is not fixed. This effect is also visualized in the mosaic plot in figure 5.24, which shows that the use of *how...like* and *what...like* constructions is correlated with the evolutionary stage of a variety: While the Phase-III data contain a considerable proportion of *how...like* tokens, the *what...like* construction increases in use across Dynamic Model stages. Furthermore, figure 5.24 illustrates that Stage V displays the highest degree of variation in the verb slot of the *what...like* construction. Speakers of the most advanced varieties do not just use *WH...like* constructions with *be* or sensory verbs such as *look*, *sound*, *feel* or *taste*, but also e.g. with *turn out* (5.43a) or *get* (5.43b). In contrast, the Stage-III data almost exclusively consist of *WH...like* tokens involving the verbs *be*, *feel* or *look*.

- (5.43) (a) *Uhm what what'd your hair turn out like* <ICE-IRE:S1A-091\$A>

- (b) *I can't eat any chocolate until New Year's, (you now **what I get like!**)* <ICE-NZ:W1B-014#114:3>

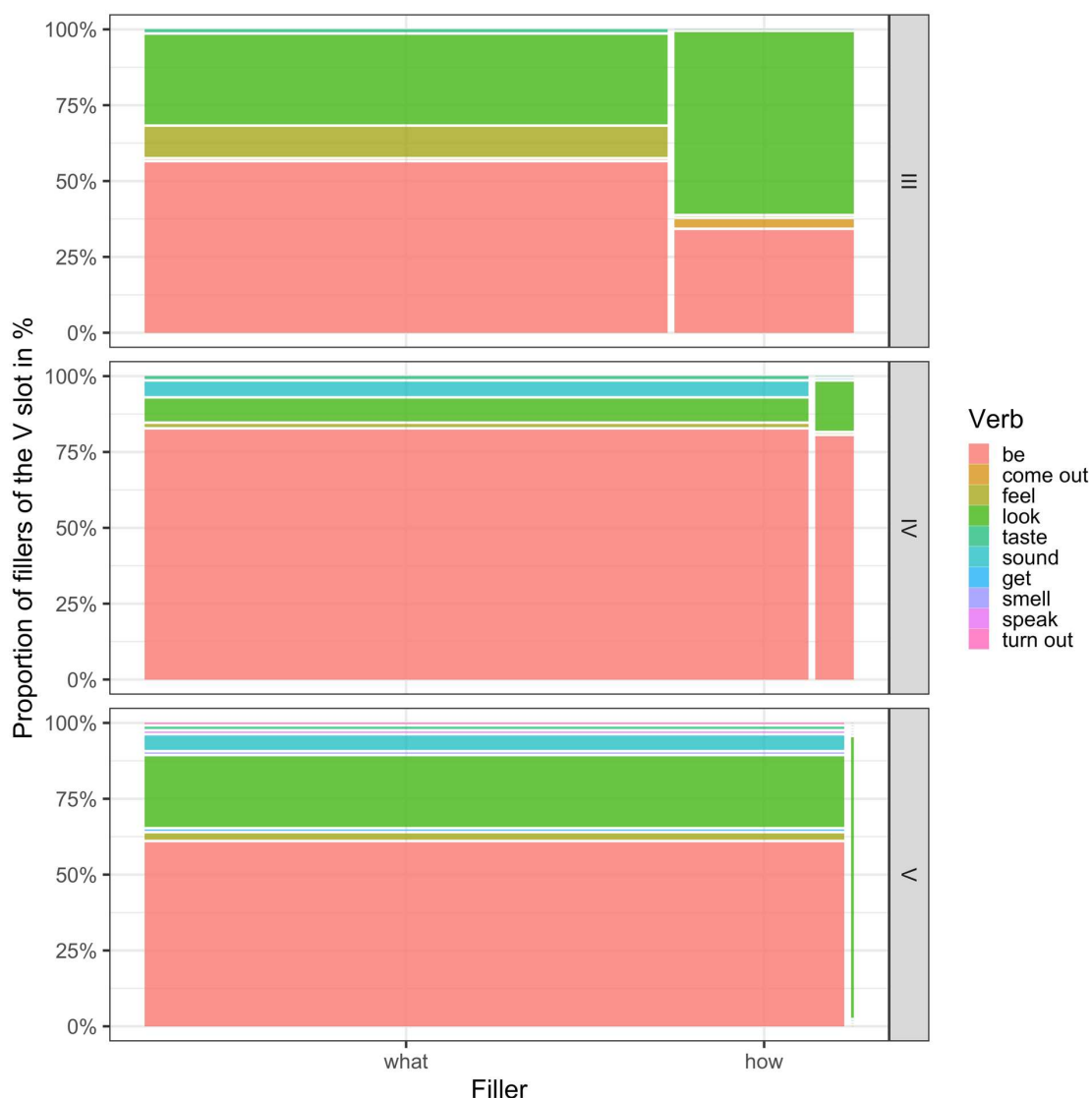


Figure 5.24: WH...V like constructions across variety phases

Since the HCFAs showed that both *how...look* and *look like* are significantly associated with Stage III, following Laporte's (2021: 332) distinction between constructions and constructs (cf. 3.3.3), the string *how...look like* (5.44a-c) should be treated as a construction that is conventionalized in the Phase-III varieties. Note that this construction would not be considered idiomatic in Standard Englishes, in which e.g. (5.44a) would be replaced by *How does the card look?* or *What does the card look like?*

- (5.44) (a) ***How's the card look like*** <ICE-HK:S1A-014#561:1:A>
- (b) *The doctor can then simulate <,> by making use of <,> computer <,> **how** the particular part would **look like** <,> after the surgery is made* <ICE-IND:S2B-027#23:1:A>

Table 5.48: HCFA results for VARIETY * FILLER * P in interrogatives (only significant types with $Q \geq 0.005$ are shown)

VARIETY	FILLER	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
CanE	where	from	28	5.4386	93.5933	>	1.921481866772E-08	***	0.01
KenE	where	from	17	2.7666	73.2269	>	2.02554558910377E-05	***	0.006
IrE	where	from	32	6.8337	92.6793	>	7.17626891911E-09	***	0.011
IrE	what	at	22	6.7324	34.6236	>	0.007284856052334	**	0.007
IrE	how	for	16	2.1473	89.3668	>	3.90283577315354E-06	***	0.006
JamE	where	from	55	6.6446	351.9015	>	3.75492025847183E-28	***	0.021
JamE	what	like	60	33.2153	21.5991	>	0.048791111914421	*	0.012
NigE	where	from	28	3.807	153.7434	>	4.16498861E-12	***	0.011
NZE	where	from	41	6.6209	178.5139	>	5.19262773658981E-16	***	0.015
NZE	how	for	15	2.0805	80.2276	>	1.95019869670797E-05	***	0.006
Phile	where	from	23	4.9184	66.4737	>	8.41461931892913E-06	***	0.008
Phile	who	with	11	0.5373	203.7374	>	5.032462797945E-08	***	0.005
SgE	where	from	19	5.415	34.0817	>	0.012884558385276	*	0.006

- (c) *but before I go into it proper let me show you **how** the henna leaf **looks like*** <ICE-NIG:dem_14:a54>

The *how...look like* construction can thus be considered a blend of a *what...like* construction with the meaning ‘how’ and a *how...look* construction (as in *How do I look?*). Apparently, Phase-III speakers prefer the literal meaning of *how* over the more idiomatic meaning of the *what...like* construction. Moreover, they seem to favor the more concrete meaning of *look*, which is semantically fully compatible with the meaning ‘how’ and frequently co-occurs with *look* (as in *how does it look?*), over the more general meaning of *be*. In accordance with Hoffmann’s (2021) predictions, the emergence of a *how...look like* construction in Phase III is therefore an example of linguistic innovation at the meso-constructional level that can be attributed to a preference for more transparent constructions in varieties at lower stages of evolution. In contrast, the single *how...look like* token that surfaces in the ICE-GB data (5.45) should be considered a construct produced by an individual speaker that is not conventionalized at the community level.

- (5.45) *That certainly is what **how** it's **looking like*** <ICE-GB:S1B-021 #138:2:B>

Besides this, the HCFA provides further support for the entrenchment of a *what...about* meso-construction in Stage III (cf. 5.3.1). As shown in table 5.46, the FILLER + P combination *what...about* occurs more frequently than expected by chance only in the Phase-III varieties, in which *what...about* makes up 22.67% (170/750) of all stranded interrogative tokens. In addition, the pattern *where...from* is entrenched in the construction networks of all variety types. Recall, however, that the HCFA of pied-piped interrogatives showed that Stage III also stores a partly schematic pied-piped *from-where* construction (cf. table 5.44). At the same time, the HCFA in section 5.3.1 indicated that Phase V has more entrenched V + *from* (e.g. *come from*, *get from*) chunks than Phase III. As *where...from* constructions frequently involve movement PPs (such as *where do you come from*), which were classified as complement_like PPs in the logistic regression analysis, these results support the assumption that the weaker stranding preference of Phase III with complement_like PPs could be due to the fact that less advanced varieties store fewer verb-preposition strings than varieties at higher Dynamic Model stages, while they have entrenched P + WH chunks that can be activated easily (cf. 5.2.3.2).

With respect to the other entrenched FILLER + P combinations, it is not surprising that *who* + P patterns turned out to be associated with both Stages V and Stages III. *Who* strongly favoring stranding (cf. 3.1.2), speakers of all variety phases should receive a lot of input for stranded constructions involving *who*. It is important to note, though, that

differences between variety phases may also be due to topic differences. For instance, the fact that *who...to* emerged as being significantly associated with Phase V only may simply be due to the fact that the ICE components of less advanced varieties might contain fewer conversations in which speakers talk about bringing or giving something to somebody.

Turning to the variety-specific HCFA results (cf. table 5.48), *where...from* occurs more frequently than expected by chance in eight varieties that represent all three Dynamic Model stages. In addition, table 5.48 shows that only the Stage-V varieties store stranded interrogative constructions for prototypical adjunct PPs (remember that e.g. movement (*where...from*) or accompaniment (*who...with*) PPs share certain properties with complements; cf. 3.2.2). Both IrE and NZE have an entrenched *how...for* construction. Moreover, IrE is significantly associated with the string *what...at*, which is typically used in constructions with time PPs (5.46a-c). Just like the results for VARIETY_TYPE, these findings thus indicate that Stage-III varieties only possess FILLER + P combinations for prototypical stranding contexts, whereas Phase V also has entrenched FILLER + P patterns for interrogative constructions with adjunct PPs, which should be dispreferred from a processing perspective.

- (5.46) (a) *What **what time**'s she coming **at*** <ICE-IRE:S1A-095\$A>
 (b) *Now **what time** is he <,> does he have woodwork **at** though* <ICE-IRE:S1B-018\$A>
 (c) ***What time** did I go in **at*** <ICE-IRE:S1B-018\$C>

However, HCFAs being conservative and unlikely to detect significant effects with small sample sizes, it must be kept in mind that the results listed in table 5.48 most likely do not include all meso-constructions that can be assumed to be conventionalized in the various varieties.

5.3.3 Fillers of the (ANTECEDENT +) P + WH slot in finite RCs

In 5.2.4.2 it was argued that speakers of all varieties should receive lots of positive input for pied-piped finite RC constructions, which can be assumed to contribute to the entrenchment of (ANTECEDENT +) P + WH chunks in the mental grammars of all speakers. This claim was explored with the help of an HCFA which tested the associations between the P slot, the WH slot and VARIETY_TYPE/VARIETY in finite RCs. The results are given in table 5.49, which lists 4 entrenched WH + P combinations for Phase V, 1 for Phase IV, and 5 for Phase III. Note, however, that these results should be taken

with a grain of salt as the frequencies (6) of the P + WH combinations *among whom* and *under whose* are very low. Consequently, these P + WH chunks cannot necessarily be assumed to be conventionalized in the Phase-III varieties. That only one significant configuration is significantly associated with Phase IV is again due to the low token numbers of the Stage-IV varieties (cf. table A.24 in the Appendix).

The most interesting finding of the HCFA concerns the P + WH combination *in which*, which is normally considered “the fronting prototype” (Jach 2021: 361) in finite RCs and was thus assumed to be deeply entrenched in the mental grammars of all speakers (cf. 3.1.2). However, as shown in table 5.49, *in which* occurs more frequently than expected by chance only in the Phase-V varieties, where it makes up almost half (41.87%, 742/1,772) of all pied-piped finite RC tokens. This effect is also illustrated in figure 5.25, which visualizes the relative frequencies of P + WH combinations across variety types. On the one hand, figure 5.25 shows that *which* is by far the most frequent filler in all variety phases, which suggests that all varieties have a partly schematic pied-piped *which*-relative clause construction in which the WH-slot is fixed. On the other hand, the stacked bar charts illustrate that Phase V has the highest proportion of *in-which* tokens, while *which* co-occurs more frequently with other prepositions in Phases III and IV. Put differently, even though *in which* is the most frequent P + WH combination across all variety types, the cue validity of *which* for *in* is lower in the less advanced varieties, which explains why the combination *in which* did not turn out to be significantly associated with the lower Dynamic Model stages.

One reason for the weaker entrenchment of *in which* in the less advanced varieties could be that *in* is frequently used in a non-literal sense in pied-piped finite RCs. For example, in the respect PP in (5.47), *in* does not have a locative meaning but it is used in an abstract sense to identify the situation in which the action described by the relative clause occurred. As L2 Englishes have a preference for transparent constructions (cf. 2.4.2), the chunk *in which* may be prone to innovation in PCEs.

(5.47) *a situation **in which** you were in danger of death* <ICE-NZ:S1A-033#1:1:B>

With respect to the other significant HCFA results, once again, the movement PP chunk *from where* emerged as being significantly associated with Phase III (5.48):

(5.48) *they belong to the two carriers Philippine Airlines and the carrier of the country **from where** they originate* <ICE-PHI:S1B-027#103:1:G>

Table 5.49: HCFA results for VARIETY_TYPE * P * WH in finite RCs (only significant types are shown)

VARIETY_TYPE	P	WH	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
V	in	which	742	624.1771	22.2409	>	0.000413815424948	***	0.028
V	for	whom	26	7.3325	47.5248	>	6.30095192032613E-05	***	0.004
V	with	whom	25	5.9097	61.668	>	3.97992112578516E-06	***	0.004
V	from	whence	3	0.059	146.6014	>	0.030507305786647	*	0.001
IV	with	whom	20	2.6881	111.4921	>	1.136229627421E-08	***	0.004
III	from	where	30	0.8852	957.6046	>	3.5781623727854E-32	***	0.006
III	with	whom	30	7.3938	69.1174	>	3.158736053588E-07	***	0.005
III	to	whom	31	12.8708	25.5359	>	0.011688075314987	*	0.004
III	among	whom	6	0.5751	51.1729	>	0.028693938296476	*	0.001
III	under	whose	6	0.3761	84.0953	>	0.002662117727024	**	0.001

Table 5.50: HCFA results for VARIETY * P * WH in finite RCs (only significant types are shown)

VARIETY	P	WH	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
CanE	with	whom	9	1.0105	63.1688	>	0.004592936438075	**	0.002
KenE	from	where	6	0.1537	222.3762	>	6.04978868305234E-05	***	0.001
TanE	to	whom	11	1.9042	43.4479	>	0.019744210476251	*	0.002
IndE	from	where	10	0.2432	391.4274	>	5.9808660713E-10	***	0.002
IndE	due_to	which	11	1.4074	65.3815	>	0.0011141629847	**	0.002
JamE	with	whom	12	1.871	54.8352	>	0.002576315864234	**	0.002
NigE	from	where	6	0.1318	261.2729	>	2.4452202352897E-05	***	0.001
PhilE	by	which	31	12.2257	28.8306	>	0.017622642940361	*	0.004
PhilE	with	whom	11	1.0639	92.7964	>	7.00865854020868E-05	***	0.002
PhilE	from	where	6	0.1274	270.702	>	2.00319153268322E-05	***	0.001
SgE	with	whom	8	0.8171	63.1429	>	0.00897682744316	**	0.001

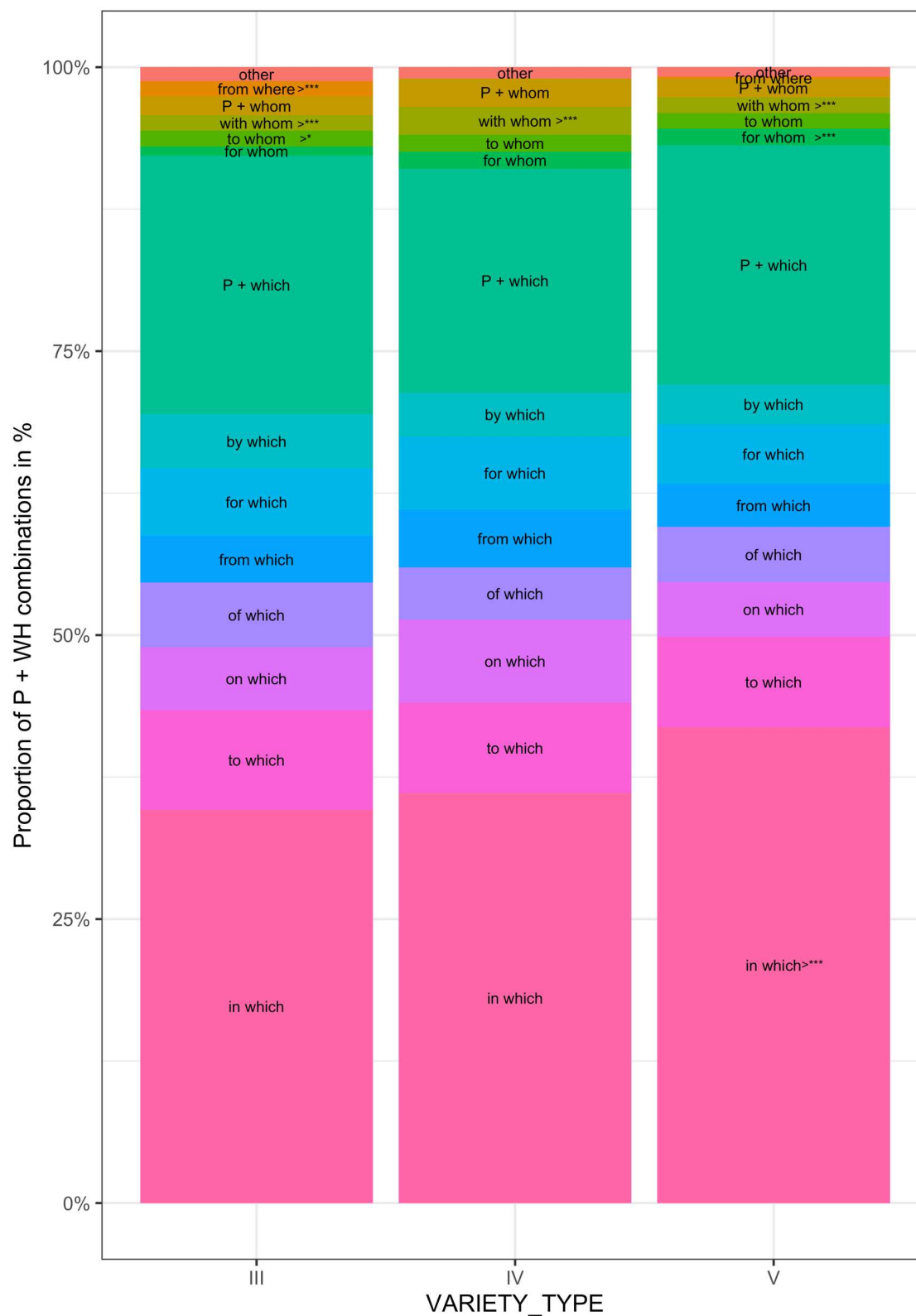


Figure 5.25: P + WH combinations in finite RCs across variety phases. Significant HCFA types are marked by >*(**).

This indicates that Phase III does not only have a pied-piped *from-where* interrogative construction (cf. 5.4.2) but also a pied-piped *from-where* relative clause construction. At the same time, there are only 3 *from-where* finite RC tokens in the Stage-

V data, which suggests that speakers of less advanced varieties overuse pied-piping constructions in contexts in which varieties at higher developmental stages would choose stranding. Considering the strong pied-piping preference of *whom* (cf. 3.1.2), it is not surprising that table 5.49 also contains a number of P + WH combinations involving *whom*. Note that the entrenchment of partly schematic pied-piped *from-where* and *with-whom* relative clause constructions in Phase III is again consistent with the assumption that Stage-III speakers show a weaker preference for stranding with complement-like PPs because they are more likely to rely on entrenched pied-piped meso-constructions than on competitors with stranded prepositions (cf. 5.2.3.2, 5.3.2). Although *with whom*, which typically appears in accompaniment PPs (e.g. *the person with whom I worked*), is entrenched in all variety phases, it can be expected that speakers of less advanced varieties generally receive less input for constructions, which means that they should possess fewer V + *with* combinations such as *work with* or *play with*, which typically favor stranding in the most advanced varieties.³²

Coming to the significant association between *among whom* and Phase III, interestingly, four of the six *among-whom* tokens in the Stage-III data involve the pattern *among-whom-BE-SUBJ* (5.49).

- (5.49) (a) *a total of ten persons among whom **among whom** were the accused* <ICE-HK:S2A-064#54:1:A>
- (b) *twelve civilians **among whom** were seven women and three children* <ICE-EA-T:S2B015T>
- (c) *great grandchildren **among whom** is Mister Yinka Ekundayo of NTA Ibadan* <ICE-NIG:bnew_31:a134>
- (d) *great journalists ; premier **among whom** (and still alive, thank God!) were Peter Enahoro and Sam Amuka-Pemu* <ICE-NIG:bl_52>

That *under whose* emerged as being significantly associated with Stage III although there are only 6 *under-whose* tokens in the entire dataset can be explained by the fact that *under whose* exclusively appears in the Phase-III varieties. One reason for this effect could be that more advanced varieties might tend to replace *under-whose* constructions such as (5.50) with structures that involve an *of*-PP (e.g. *under the tenure*

³² That the HCFAs in 5.3.1 did not identify cross-varietal differences regarding the entrenchment of V + *with* patterns might be due to the fact that HCFAs are conservative with small sample sizes. Consequently, it is likely that they did not identify all V + P combinations that can be assumed to be entrenched in the individual varieties/variety types.

of whom; Huddleston, Pullum & Peterson 2002: 1049). However, as the low token numbers in the present analysis do not allow a thorough investigation of these phenomena, the use of *among whom* and *under whose* should be investigated in more detail in future studies using larger corpora.

(5.50) *It said former President Olusegun Obasanjo and former Head of State General Abdulsalami Abubakar (rtd) **under whose tenure** the scam was perpetrated should be interrogated by the panel <ICE-NIG:Pr_31>*

Turning to the variety-specific results, the low frequencies of the P + WH combinations listed in table 5.50 should also be taken with caution. Nevertheless, it is interesting that there is only one P + WH chunk that is significantly associated with Phase V (*with whom* in CanE). As the most advanced varieties strongly favor the chunk *in which*, the frequencies of other P + WH combinations might be too low to emerge as significant in the HCFA. Moreover, these results again show that a partly schematic *from-where* construction can be said to be entrenched in several Stage-III varieties, namely in KenE, IndE, NigE and PhilE.

Next, an HCFA was performed which tested the associations between the ANTECEDENT slot, the P + WH slot and VARIETY_TYPE/VARIETY. However, as some of the factor associations in table 5.51 are based on very low frequencies, once again, the results have to be taken with a grain of salt and cannot necessarily be considered indicators of collective entrenchment.

While Stages III and V turned out to be significantly associated with 11 ANTECEDENT + P + WH combinations each, the HCFA identified 3 combinations that occur more frequently than expected by chance in Stage IV (cf. table 5.51). Disregarding the results for Phase IV, which are based on low token numbers, there are thus no clear quantitative differences between higher and lower Dynamic Model stages. Furthermore, all variety types have entrenched ANTECEDENT + P + WH strings for respect (*level at which, circumstance under which, point at which, framework within which*), manner (*manner in which, way in which, ease with which*), and degree (*extent to which, degree to which, rate at which*) PPs, which can be explained by the almost categorical pied-piping preference of these adjunct PPs. However, with respect to manner PPs, it is important to note that *way in which* is significantly associated with Phase V only, while *manner in which* turned out to be entrenched in Phase III. Moreover, the string *way in which* accounts for 8.45% (144/1,705) of all pied-piped finite RC tokens in Stage V, whereas *manner in which* only makes up 1.86% (38/2,048) of the pied-piped finite RC data in Phase III.

Table 5.51: HCFA results for VARIETY_TYPE * ANTECEDENT * P_WH in finite RCs (only significant types are shown)

VARIETY_TYPE	ANTECEDENT	P_WH	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
V	way	in_which	144	43.4838	232.351	>	2.90261611047992E-28	***	0.022
V	extent	to_which	31	2.8141	282.3087	>	2.01482949381514E-16	***	0.006
V	degree	to_which	19	1.0318	312.9058	>	1.74863856E-12	***	0.004
V	point	at_which	13	0.6188	247.728	>	5.60674264189E-08	***	0.003
V	rate	at_which	15	0.4678	451.4426	>	1.75539869E-12	***	0.003
V	means	by_which	8	0.4436	128.7177	>	0.008056863091149	**	0.002
V	method	by_which	7	0.3644	120.832	>	0.039586996575501	*	0.001
V	period	during_which	7	0.2586	175.7404	>	0.003941183957643	**	0.001
V	person	to_whom	6	0.1761	192.6054	>	0.011468271436057	*	0.001
V	ease	with_which	6	0.2206	151.4119	>	0.042694613306442	*	0.001
V	framework	within_which	5	0.0675	360.4379	>	0.00356700334992	**	0.001
IV	extent	to_which	20	1.2692	276.4284	>	4.50314473E-12	***	0.004
IV	purpose	for_which	8	0.2274	265.6698	>	4.65984518466073E-05	***	0.002
IV	means	by_which	6	0.2001	168.1101	>	0.024161997606817	*	0.001

Table 5.51 (cont.)

VARIETY_TYPE	ANTECEDENT	P_WH	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
III	extent	to_which	37	3.3802	334.3858	>	2.88818848012286E-20	***	0.007
III	manner	in_which	38	13.0579	47.6423	>	0.0046545977145	**	0.006
III	purpose	for_which	12	0.6057	214.3472	>	9.2897474593756E-07	***	0.003
III	level	at_which	11	0.3988	281.809	>	2.2539969224046E-07	***	0.002
III	rate	at_which	9	0.562	126.6901	>	0.002982679312692	**	0.002
III	means	by_which	11	0.5328	205.6349	>	4.82841128004587E-06	***	0.002
III	process	by_which	10	0.9134	90.3945	>	0.015598083047961	*	0.002
III	person	to_whom	7	0.2115	217.89	>	0.001004741472798	**	0.002
III	circumstance	under_which	11	0.3606	313.9125	>	7.6970751019E-08	***	0.002
III	place	from_where	4	0.0279	565.5046	>	0.008011321489369	**	0.001
III	ease	with_which	7	0.265	171.1707	>	0.004646174645154	**	0.001

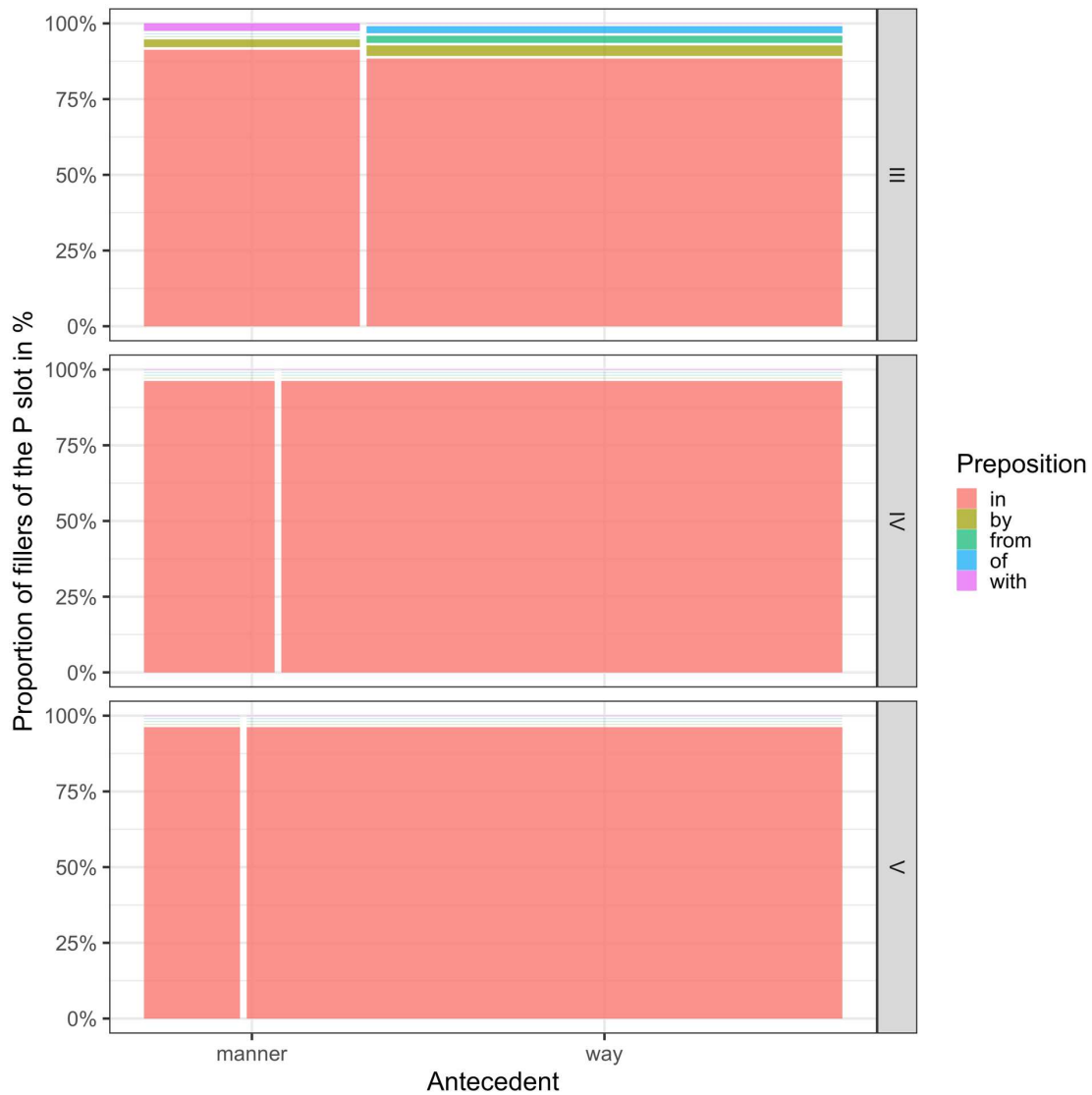


Figure 5.26: way/manner-P-which across variety phases

Even though less advanced varieties are thus characterized by a higher proportion of *manner-in-which* and a lower proportion of *way-in-which* tokens, this does not mean that Stages III and IV do not use *way-in-which* constructions. Instead, the cue validity of *way* for *in which* is simply weaker in PCEs than in Standard Englishes: As shown by the mosaic plot in figure 5.26, *way* is always followed by *in which* in Stage V, while Phase III exhibits more variation in the P slot. Examples (5.51a-c) illustrate that Stage-III speakers do not only use the chunk *way in which* but also e.g. *way by which* (5.51a), *way from which* (5.51b) or *way of which* (5.51c) (also see table 5.52).

- (5.51) (a) *the allowed **ways by which** these sounds can be combined* <ICE-PHI:W1A-019#91:2>
- (b) *The **ways from which** he can obtain labour force* <ICE-EA-T:S1B004T>

- (c) *the way of which* words are governed and used under grammatical rules <ICE-EA-T:W1A001T>

Table 5.52: *way/manner-P-which* across variety phases (absolute and relative frequencies)

VARIETY_TYPE	ANTECEDENT	P	Abs. freq.	Rel. freq.
III	manner	by	1	2.56%
		in	37	94.87%
		with	1	2.56%
	way	by	3	3.49%
		from	2	2.33%
		in	79	91.86%
		of	2	2.33%
	manner	in	10	100%
IV	way	in	43	100%
V	manner	in	23	100%
	way	in	143	100%

On the one hand, the higher proportion of *manner-in-which* constructions in PCEs can be considered a quantitative innovation. On the other hand, tokens such as *way of which* or *way from which* show that qualitative innovations in PCEs typically emerge the meso-constructural level (cf. 2.2.3.1). Since speakers of varieties at earlier stages receive less input, their construction networks are not as deeply entrenched as those of more advanced varieties. Consequently, particularly partly schematic constructions that involve both substantive and schematic slots (*way – P – which*) are likely to be affected by innovation. The results regarding *way in which* thus confirm Laporte’s (2021: 321) claim that, “due to less entrenched constructions in less advanced varieties, more variation is found in their instantiations”. Moreover, the weaker entrenchment of *way in which* in the less advanced varieties is compatible with the finding that the string *in which* is not as deeply entrenched in earlier Dynamic Model stages. Importantly, though, tokens such as (5.51a-c) do not represent conventionalized constructions of the Stage-III varieties as the token numbers of these structures are very low (cf. table 5.52).

Another HCFA result that shows that the string *in which* is less deeply entrenched in varieties at lower developmental stages is the significant association between *circumstance under which* and Phase III (5.52). While there are 11 *circumstance-under-which* tokens in the Stage-III data, only 2 such cases occur in the Stage-V varieties. At the same time, the most advanced varieties have 9 *circumstance-in-which* tokens (5.53).

That *circumstance in which* did not turn out to be significantly associated with Phase V can be explained by the fact that the Stage-V data includes so many ANTECEDENT + *in which* types that the token numbers of *circumstance in which* are too low to emerge as significant. Interestingly, however, in all of the Phase-III examples in (5.52) *under which* can be replaced by *in which*.

- (5.52) (a) *the **circumstances under which** obtaining a written consent of the owner is necessary* <ICE-HK:W2A-020#64:1>
 (b) *the **circumstances under which** bridewealth can be returned or may not be returned* <ICE-EA-T:W2B016T>
 (c) *The **circumstances under which** the Thomasites worked* <ICE-PHI:W2A-001#32:1>
 (d) *the **circumstances under which** audit reports are qualified to gain insight into the causes and extent of qualified reports* <ICE-IND:W2A-020#72:1>
 (e) *the **circumstances under which** Ms. Imire died* <ICE-EA-K:S1BHN07K>
- (5.53) *All sought, at minimum, the advance publication of legislation that would precisely identify **circumstances in which** abortions could be performed* <ICE-IRE:W2A-020\$B>

Now, in its literal sense, *under* is used to express vertical direction. However, through metaphorical extension, *under* can also assume the meaning ‘subjection’ or ‘subordination’ (Quirk et al. 1985: 685–686), which is highly compatible with the concept ‘passive subordination’ that is typically expressed by the string *circumstances under which*. Varieties at earlier stages may thus prefer *circumstances under which* over *circumstances in which* as the former variant is more transparent. Such an explanation is in line with the idea that *in which* might be dispreferred in earlier stages of Schneider’s model as it is frequently used in a non-literal sense. Moreover, in accordance with Rohdenburg’s (1996: 151) Complexity principle (cf. 3.3.3), speakers of PCEs may prefer more explicit variants such as *under which* over *in which* due to the high processing complexity of finite RCs.

Besides this, the string *person to whom* typically appears in constructions with optional complement PPs (5.54a) and prepositional verbs (5.54b), which strongly favor pied-piping in other clause types (cf. 5.2.2.2, 5.2.3.2). The entrenchment of *person to whom* in both Stages III and V thus supports the claim made in 5.2.4.2 that speakers may favor pied-piping with finite RCs regardless of PP types due to the entrenchment of partly schematic pied-piping constructions which can be activated easily.

- (5.54) (a) *The speaker values the person **to whom** he is talking more than at any other time* <ICE-IRE:W1A-012\$A>
- (b) *these kinship relationships are still expressed in honour of the people **to whom** the house belongs* <ICE-NZ:W2A-004#45:1>

In addition, the HCFA identified a *place-from-where* meso-construction for Phase III (5.55):

- (5.55) (a) ***The place from where** the accused took out uh the suitcase* <ICE-IND:S2A-064#57:1:A>

Although there are only 4 *place-from-where* tokens in the data, this corroborates the assumption that Stage-III favors pied-piping with movement PPs due to the entrenchment of pied-piped meso-constructions that are preferred over constructions with stranded prepositions (cf. 5.3.2).

Coming to the variety-specific HCFA results given in table 5.53, once more, *way in which* turned out to be significantly associated with Stage-V varieties only, namely with BrE, IrE and NZE. Although the frequencies of the other significant types are very low, interestingly, the entrenched combinations of BrE, CanE, IrE and NZE only involve manner, degree and respect adjunct PPs, which have an almost categorical pied-piping preference. At the same time, the chunks that are significantly associated with the less advanced varieties include a wider range of adjuncts such as location (*road on which*) or instrument (*means by which*) PPs.

In sum, the HCFAs did not identify clear quantitative differences between variety phases with regard to the entrenchment of (ANTECEDENT +) P + WH combinations. However, the findings suggest that certain meso-constructions that are deeply entrenched in the Phase-V varieties, such as pied-piped *in-which* and *way-in-which* relative clause constructions, are not as deeply entrenched in varieties at earlier stages of evolution. The construction networks of less advanced varieties being less deeply entrenched, these speakers show more variation in the preposition slot of pied-piped finite WH-relative clause constructions.

Table 5.53: HCFA results for VARIETY * ANTECEDENT * P_WH in finite RCs (only significant types are shown)

VARIETY	ANTECEDENT	P_WH	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
CanE	degree	to_which	8	0.1797	340.3288	>	2.95730227984173E-05	***	0.002
CanE	point	at_which	6	0.1078	322.0596	>	0.002556921513767	**	0.001
TanE	extent	to_which	14	0.5117	355.5486	>	7.6054304943E-10	***	0.003
BrE	way	in_which	57	16.8069	96.1204	>	1.341638518075E-08	***	0.009
BrE	rate	at_which	11	0.1808	647.4286	>	1.8317853953E-10	***	0.002
BrE	extent	to_which	12	1.0877	109.4771	>	0.0026847128443	**	0.002
IndE	road	on_which	6	0.0639	551.4442	>	0.000115217867791	***	0.001
IrE	way	in_which	38	10.4565	72.5524	>	4.85960493886491E-05	***	0.006
JamE	extent	to_which	18	0.8913	328.4053	>	1.0602417E-11	***	0.004
JamE	means	by_which	6	0.1405	244.3683	>	0.012192817492015	*	0.001
NigE	tenure	during_which	3	0.0029	3097.4512	>	0.0052621900493	**	0.001
NZE	way	in_which	38	8.6458	99.6633	>	2.0129996576996E-07	***	0.007
SgE	purpose	for_which	6	0.0677	519.8255	>	0.000162912819394	***	0.001

5.3.4 The P + WH slot in free RCs

Next, a closer look was taken at all P + FILLER combinations in free RCs. Although pied-piped free RC tokens occur in all varieties except CanE, the absolute frequencies are higher in varieties at earlier stages of Schneider's cycle (cf. table 5.54). While the low frequencies did not permit an HCFA, a qualitative analysis of these tokens shows that pied-piping in free RCs mainly occurs with P + FILLER chunks that are likely to be stored as lexical items. As shown in table 5.55, the three P + WH combinations *in wh-ever*, *from wh-ever* and *to what* account for more than half (50.94%) of all pied-piped free RC tokens in the dataset.

Table 5.54: Pied-piped free RC tokens across varieties (absolute frequencies)

VARIETY (VARIETY_TYPE)	Abs. freq.
IndE (III)	12
KenE (III)	6
TanE (III)	6
JamE (IV)	6
PhilE (III)	6
NigE (III)	5
BrE (V)	4
HKE (III)	2
IrE (V)	2
NZE (V)	2
SgE (IV)	2
CanE (V)	0
Total	53

While 9 of the *in-wh-ever* tokens involve the complex filler phrase *in whichever/whatever way* (5.56), the frequent occurrence of *to what* can be explained by the chunk *to what extent* (5.57).

(5.56) ***In whichever way*** you pronounce it <,,> the accent must be on the second syllable <ICE-IND:S2B-026#110:1:A>

(5.57) That is ***to what extent*** individuals know both language <ICE-EA-T:W1A011T>

Table 5.55: *P + WH combinations in free RCs (absolute and relative frequencies)*

P + WH	Abs. freq.	Rel. freq.
in wh-ever	15	28.30%
from wh-ever	6	11.32%
to what	6	11.32%
from where	5	9.43%
on what	3	5.66%
to whom	3	5.66%
on whose	2	3.77%
to which	2	3.77%
at wh-ever	1	1.89%
by wh-ever	1	1.89%
by what	1	1.89%
from whom	1	1.89%
from whose	1	1.89%
in what	1	1.89%
in which	1	1.89%
on wh-ever	1	1.89%
on which	1	1.89%
through wh-ever	1	1.89%
with whom	1	1.89%
Total	53	100%

Now, instances of *in-wh-ever-way* and *to-what-extent* can be found in all variety phases. In contrast, *from-wh-ever* and *from-where* tokens almost exclusively occur in PCEs (there is a single *from-whenever* token in IrE). Note that examples (5.58a-b), in which the free RCs function as optional adverbials, provide support for Hoffmann's (2011: 262) prediction that pied-piping in free RCs should be possible if the form of the filler of the free RC is compatible with the requirements of the verb of the main clause. Interestingly, however, in (5.58a) the pied-piped variant is used although the construction is a free RC involving an obligatory complement with *be*, i.e. a construction type which would strongly favor stranding in Stage V (cf. 5.2.2.2). Remember that *from where* also turned out to be significantly associated with Phase III in the HCFAs of pied-piped interrogatives and finite RCs (cf. 5.3.2 and 5.3.3). Speakers of less advanced varieties may thus favor pied-piped free RC constructions with *from-wh(ever)* since the entrenched chunk *from-where* has a high chance of being activated. At the same time, in example (5.59), in which *on which* could be paraphrased by *the topic on which*, the subject position

of the matrix clause would normally have to be filled by an NP but the filler phrase of the free RC is a PP (*on which*). This shows that PCEs license pied-piping in free RCs even when the form of the free RC filler is not compatible with the requirements of the main verb of the matrix clause.

- (5.58) (a) *we need to encourage the airlines of all kinds all charters whoever wants to come to Tanzania it's open-door open-skies welcome to Tanzania Karibu Tanzania **from wherever** you are* <ICE-EA-T:S2B021T>
- (b) *IN THE BEGINNING, he was dust again, except that which was left of his mortal self was contained in a white marble vessel – smooth, symmetrical **from whatever angle** one viewed it* <ICE-PHI:W2F-009#3:1>
- (5.59) ***on which** I am presenting a paper is the theme of voyage <,,> theme of flight* <ICE-IND:S2A-022#14:1:A>

Although no statistical analysis of the pied-piped free RC tokens was possible, the study corroborates Hoffmann's (2011) finding that pied-piping in free RCs is not impossible, particularly with structures involving lexicalized manner and degree PP chunks such as *in whichever way* or *to what extent*. While pied-piping in free RCs is a phenomenon that must definitely be investigated using larger sample sizes and multifactorial statistical analysis, one reason for the fact that most of the pied-piped free RC tokens in the present study surfaced in the PCE data could be that these speakers generally receive more input for pied-piped prepositions in adjunct PPs (cf. 5.2.2.2, 5.2.3.2). As a result, speakers of less advanced varieties may also be likely to use entrenched P + WH strings in free RCs.

5.3.5 Summary

Regarding hypothesis IV, which assumes that more advanced varieties should have more entrenched meso-constructions, while less advanced varieties should rely more on prototypical and frequent fillers, the HCFAs presented in this section provided mixed results. Hypothesis IV was only confirmed for non-prototypical stranding environments such as stranded finite RC or stranded adjunct interrogative constructions:

In stranded finite RC constructions, Phase III can be said to rely more on entrenched V + P combinations than Phase V (cf. 5.3.1).

With respect to stranded interrogative constructions, Phase III only stores prototypical FILLER + P meso-constructions (e.g. *what...about, where...from*),

whereas Stage V also has FILLER + P combinations for constructions with adjunct PPs (e.g. *how...for*). Moreover, the less advanced varieties seem to rely on more specific *what...talk about* and *how...look like* constructions in which the V slot is fixed, whereas Phase V has more abstract *what...about* and *what...like* constructions (cf. 5.3.2).

In contrast, in finite RCs, i.e. in prototypical pied-piping environments, it is Phase V which turned out to rely more on entrenched meso-constructions:

While a partly schematic pied-piped *in-which* finite relative clause construction is deeply entrenched in the construction network of Stage V, Phase III shows more variation in the preposition slot of pied-piped finite RCs (cf. 5.3.3).

More advanced varieties also possess a deeply entrenched pied-piped *way-in-which* construction. In contrast, Phase III is not only more likely to use the string *manner in which*, but it also exhibits more variation in the P slot of a *way-P-which* construction (cf. 5.3.3).

At the same time, Phase III has partly schematic *from-where* constructions for pied-piped interrogatives, finite RCs and free RCs. As *from where* typically appears in movement PPs, i.e. a PP type associated with stranding in Phase V, this shows that speakers of less advanced varieties overuse pied-piped prepositions in contexts in which Standard varieties would favor stranding. This can be explained by the fact that speakers of varieties at earlier stages generally receive more input for pied-piped than for stranded prepositions (e.g. with adjunct PPs and in more formal contexts; cf. 5.2.2.2, 5.2.4.2).

While the HCFAs identified entrenched meso-constructions that are significantly associated with individual varieties or variety types, in the next section LNRE models will be fitted in order to compare the productivity of constructional slots across variety phases.

5.4 The productivity of constructional slots

The LNRE models presented in this section aim to test hypothesis V, which predicts a correlation between the productivity of preposition placement constructions and Dynamic Model phases. LNRE models were fitted to investigate the productivity of the V + P slot in stranded constructions (cf. 5.4.1), of the FILLER + P slot in stranded free RCs and interrogatives (cf. 5.4.2), and of the ANTECEDENT slot and the P + WH slot in pied-piped finite RCs (cf. 5.4.3).

5.4.1 The V + P slot

As verbs and stranded prepositions tend to be strongly associated with each other, the first series of LNRE models analyzed the productivity of the combined V + P slot³³ in Phases III, IV and V as well as BrE. According to a chi-squared (cf. table 5.56), all these models have a reasonable goodness of fit.

Table 5.56: χ^2 - and p-values of the LNRE models of the V + P slot

VARIETY(_TYPE)	χ^2	<i>p</i>
III	2.76801439470345	0.597366513126189
IV	5.62952574541667	0.131093556315344
V	2.63491011932642	0.620651570181775
BrE	1.06980429517905	0.784367987386636

The vocabulary growth curves generated from these models (cf. figure 5.27) show that the V + P slot is most productive in Phase V, which indicates that more advanced speakers have more abstract representations of stranding constructions than speakers of less advanced varieties. At the same time, there is no difference in productivity between Stages III and IV as the confidence intervals of these curves overlap. Although the LNRE analysis does not confirm a direct correlation between the productivity of the V + P slot and Schneider's model stages, the results generally provide support for Hoffmann's (2021: 21) Dynamic Model Productivity hypothesis. Surprisingly, the productivity of the V + P slot in BrE is much lower than the average productivity of this slot in Stage V. Even though LNRE models take into account differences in sample size, this effect might be due to the low token numbers of BrE, which has a considerably smaller sample size than the other groups (cf. table A.25 in the Appendix).

³³ As in 5.3.1, the term "V + P" also includes adjective-preposition combinations.

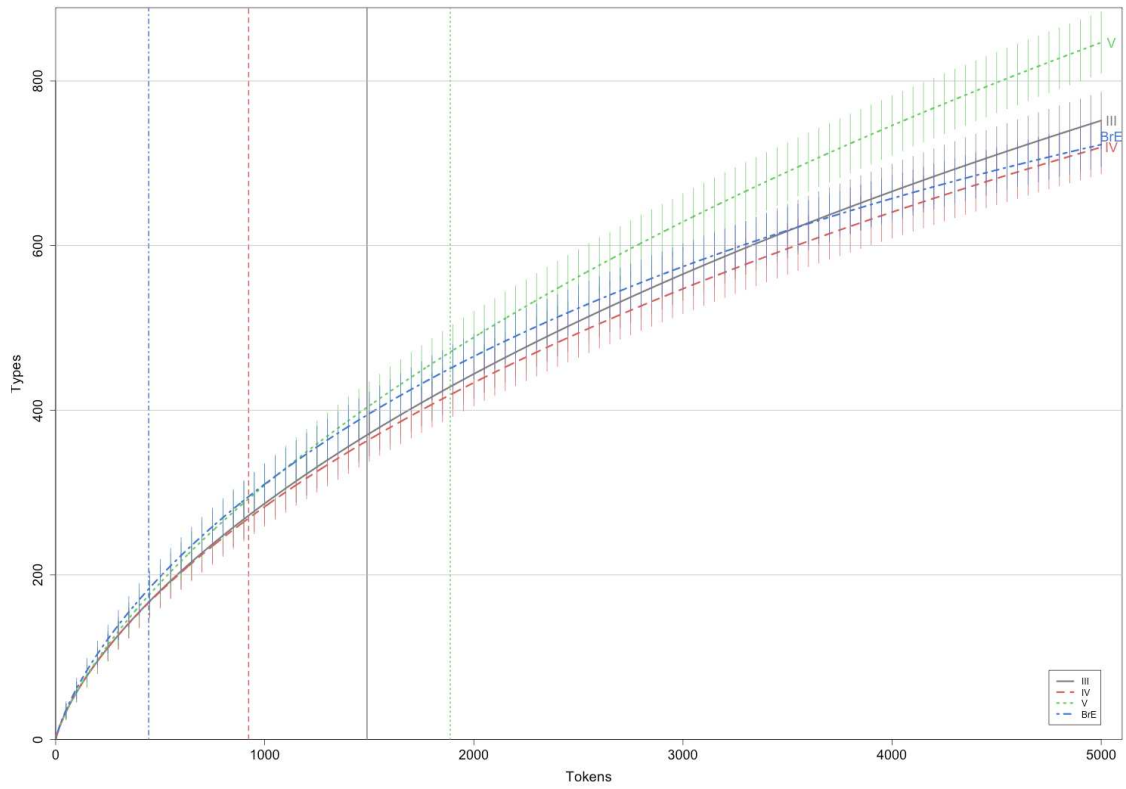


Figure 5.27: The productivity of the $V + P$ slot (LNRE models)

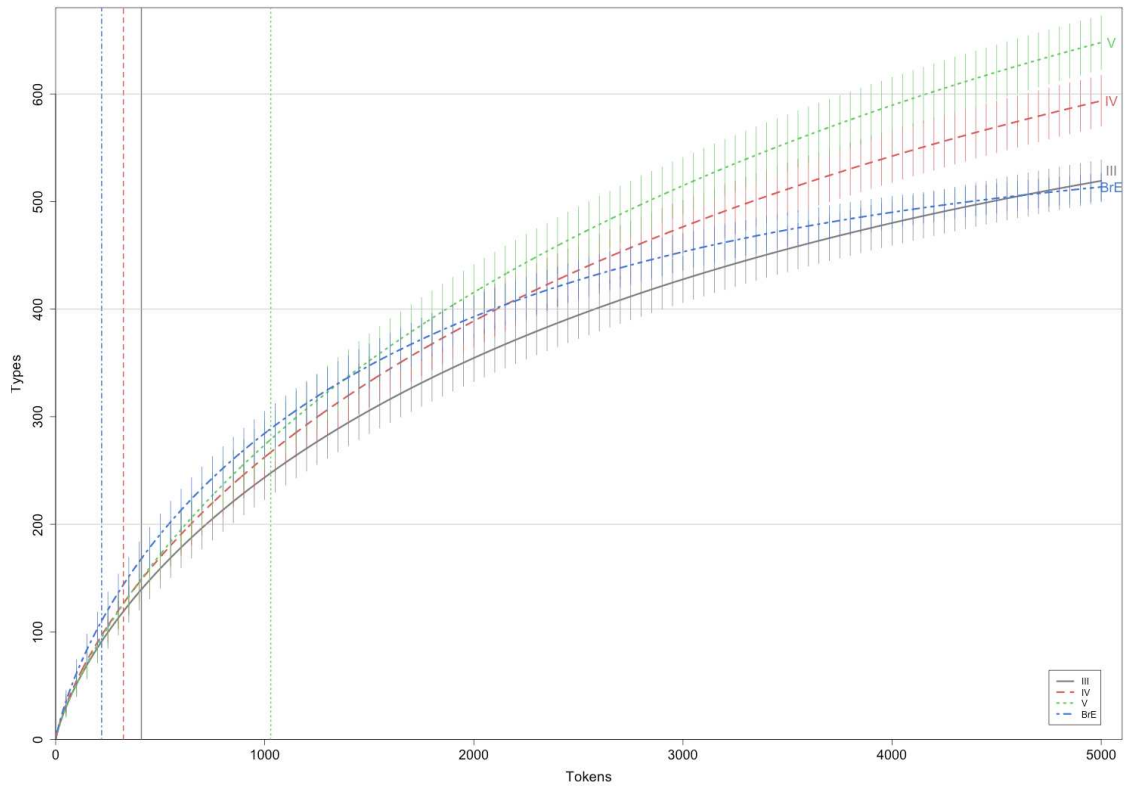


Figure 5.28: The productivity of the $V + P$ slot in informal/medium texts about everyday topics (LNRE models)

While these LNRE results indicate that the V + P slot tends to be more productive in more advanced varieties, it is likely that the curves in figure 5.27 are skewed by topic and register effects. The productivity of constructional slots may, for example, be lower in spontaneous types of discourse, in which speakers should be most likely to rely on entrenched constructions. As Schmid (2020: 76) pointed out,

[i]f entrenchment is regarded as being conducive to ease and speed of activation [...], and if spontaneous language use is subject to time pressure and limitations concerning planning, then those patterns of association which are more entrenched than others should be expected to make it to the surface in spontaneous language use without too much interference from additional linguistic or metalinguistic reflexion.

A second set of LNRE models was thus fitted which only included data from more spontaneous contexts. As data sparsity made it impossible to conduct an LNRE analysis that only included tokens from informal texts about personal topics, which would have represented the most spontaneous type of speech, it was decided to analyze the productivity of the V + P slot using all stranded tokens from texts about everyday topics (i.e. personal, situational description, sports, household and gardening) which had been categorized as informal and medium in terms of formality (cf. table 5.13). As shown in table 5.57, the results of the chi-squared test again indicate a good model fit for all variety phases and BrE. The productivity curves predicted by these LNRE models are given in figure 5.28.

Table 5.57: χ^2 - and p -values of the LNRE models of the V + P slot in informal/medium texts about everyday topics

VARIETY(_TYPE)	χ^2	p
III	3.84421901578824	0.278784444138279
IV	3.94215456288062	0.267778520222173
V	4.81638087758296	0.18574709576318
BrE	2.78447357271961	0.426062225317298

A comparison of figures 5.27 and 5.28 reveals that the overall slot productivity is indeed lower in more informal and more spontaneous text types. While the Phase-V curve in figure 5.27 predicts that the most advanced variety type exhibits more than 800 V + P types once 5,000 tokens are reached, the Stage-V curve in figure 5.28 reaches only about 630 types. As the same effect also applies to the other variety stages, this suggests that speakers generally rely more on entrenched V + P combinations in more spontaneous situations. In addition, figure 5.28 shows a direct correlation between the productivity of the V + P slot and Dynamic Model stages, with Phase III displaying the lowest and Phase V displaying the highest productivity. The only exception is again BrE, whose low

productivity may again be due to the small sample size of this variety (cf. table A.26 in the Appendix). Particularly in more informal and more spontaneous contexts speakers of less advanced varieties thus show a greater reliance on entrenched verb-preposition strings than speakers of more advanced varieties.

5.4.2 *The FILLER + P slot in free RCs and interrogatives*

Next, two LNRE analyses were carried out to investigate the productivity of the combined FILLER + P slot in stranded free RCs and interrogatives. Since free RCs are frequently considered a categorical stranding environment (cf. 3.1.1.2), separate LNRE models were fitted for stranded free RCs and for stranded interrogatives in order to explore if there is a difference in productivity between the two construction types.

Table 5.58: X^2 - and p -values of the LNRE models of the FILLER + P slot in interrogatives

VARIETY(_TYPE)	X^2	p
III	3.86248085358922	0.276701727965451
IV	4.22491662465688	0.238179068265576
V	2.16363948040184	0.539145722246037
BrE	3.65590571801627	0.301097448175121

Table 5.59: X^2 - and p -values of the LNRE models of the FILLER + P slot in free RCs

VARIETY(_TYPE)	X^2	p
III	3.68290395513564	0.29780329790373
IV	1.87666824199364	0.598394825446067
V	3.66373064873105	0.30013936987021
BrE	2.43910298245059	0.486397124483488

Beginning with the interrogative data, the productivity curves in figure 5.29 suggest that the combined FILLER + P slot in stranded interrogatives is generally not very productive (e.g. Phase V shows only about 110 types once 5,000 tokens are reached), which can be explained by the fact that the number of preposition and filler types that exist in English is much more limited than e.g. the number of verb types. Nevertheless, the curves suggest a correlation between variety types and slot productivity: While Phase V emerged as the most productive variety, the FILLER + P slot is least productive in Phase III. Stage IV displays a higher productivity than Stage III but a lower productivity than Stage V. The low productivity of BrE can again be attributed to the extremely small sample size of BrE (cf. table A.27 in the Appendix). The X^2 - and p -values in table 5.58 suggest a good fit for all four models on which these curves are based.

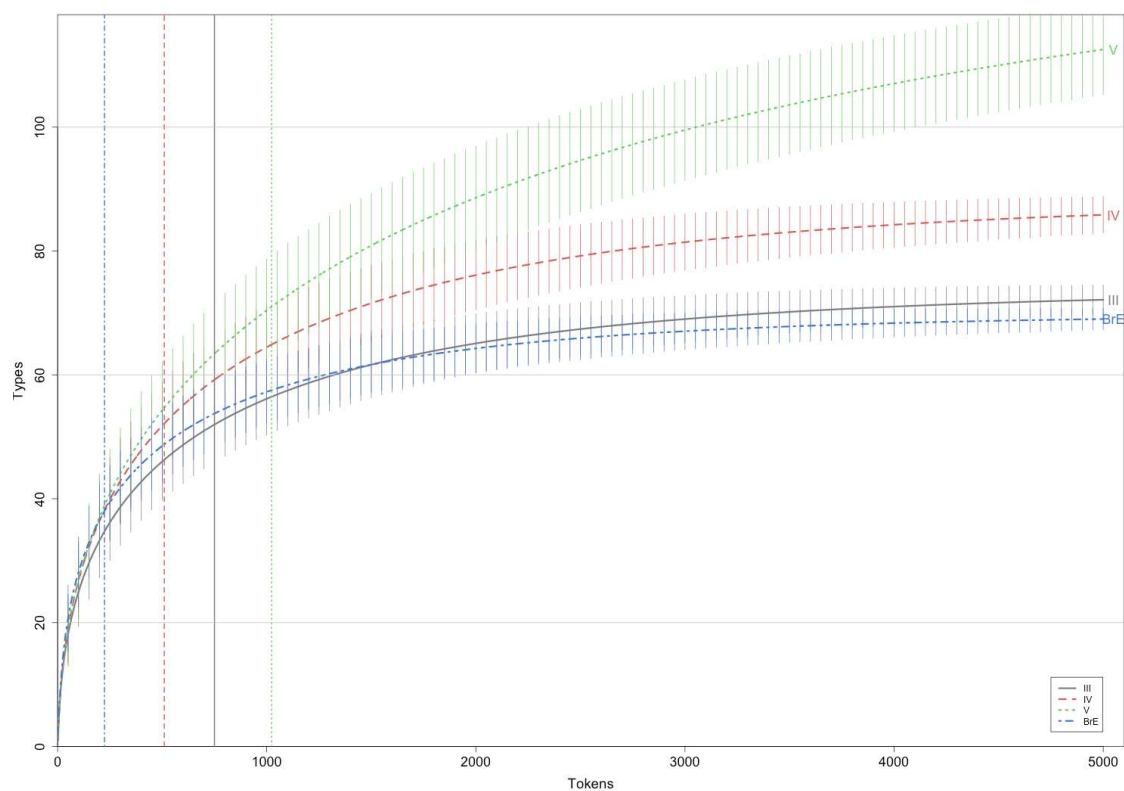


Figure 5.29: The productivity of the FILLER + P slot in interrogatives (LNRE models)

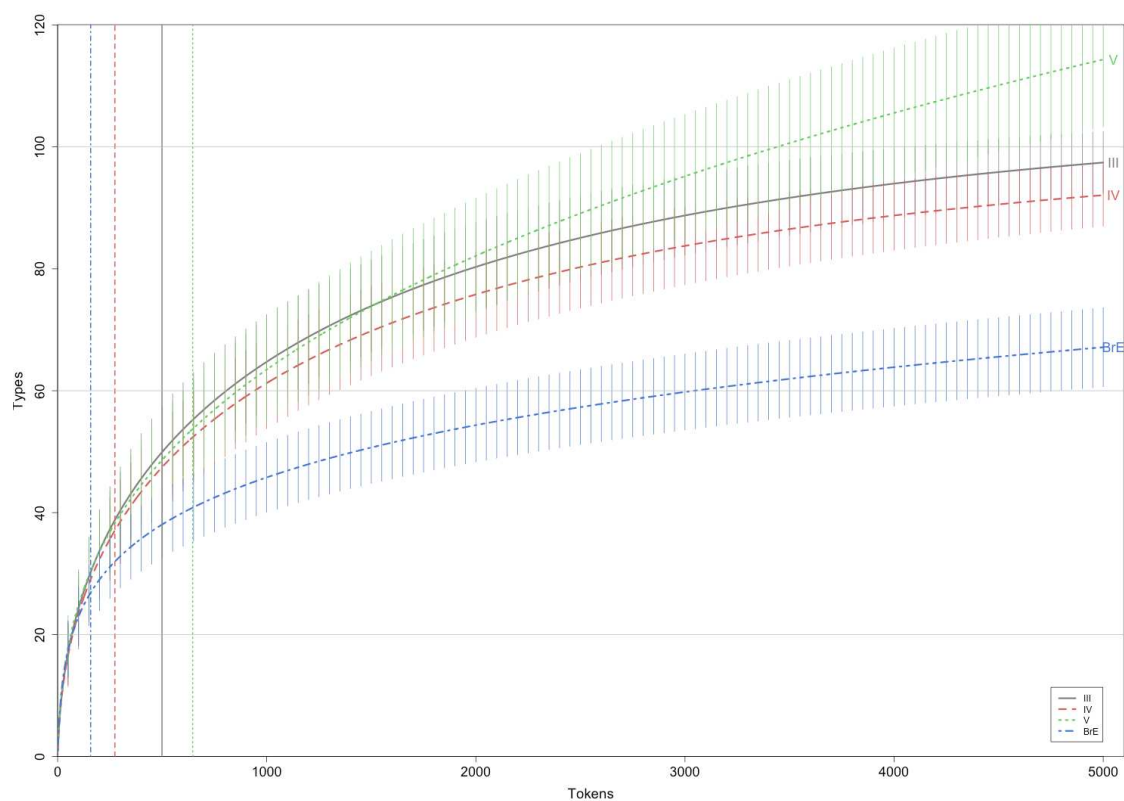


Figure 5.30: The productivity of the FILLER + P slot in free RCs (LNRE models)

That the FILLER + P slot is most productive in the most advanced varieties is also reflected in figure 5.30, which visualizes the productivity of the FILLER + P slot in stranded free RCs. While speakers of the most advanced varieties can be said to have the most abstract constructional template for FILLER + P patterns in free RCs, there is no significant difference between the curves of Phases III and IV, whose confidence intervals overlap. In addition, the slot productivity of Stage III is clearly lower in interrogatives than in free RCs (compare figures 5.29 and 5.30), which is consistent with the finding that speakers of less advanced varieties show the strongest preference for pied-piped (adjunct) interrogative constructions (cf. 5.3.2). Moreover, the difference in productivity between more and less advanced varieties is not as strong in free RCs as in interrogatives, which indicates that free RCs are very strongly associated with stranding even in lower Dynamic Model stages. That BrE again has a much lower slot productivity than all other groups can be explained by the fact that the model for BrE is based on only 158 tokens (cf. table A.28 in the Appendix). The chi-squared test results indicate a reasonable goodness of fit for all models (cf. table 5.59).

5.4.3 The ANTECEDENT and P + WH slots in finite RCs

Table 5.60: X^2 - and p -values of the LNRE models of the ANTECEDENT slot in pied-piped finite RCs

VARIETY(_TYPE)	X^2	p
III	4.10467018140651	0.662514038486369
IV	0.408005744973988	0.938583775100996
V	1.47123033521512	0.916365756665443
BrE	6.09676858256445	0.106995972460998

The final series of LNRE models analyzed the productivity of the ANTECEDENT slot and the combined P + WH slot in pied-piped finite RCs. With respect to the productivity of the ANTECEDENT slot, the curves in figure 5.31 show that there is no difference in productivity between Phases III, V and BrE. At the same time, Stage IV exhibits a slightly lower slot productivity than the other groups, which corresponds to the results of the logistic regression analysis which indicated that Phase IV has a slightly stronger preference for stranded finite RCs than Phase V (cf. 5.2.4.2). According to a chi-squared test, all these LNRE models have a reasonable goodness of fit (cf. table 5.60). While the productivity of the slots of stranding constructions turned out to be correlated with

Dynamic Model stages (cf. 5.4.1 and 5.4.2), the ANTECEDENT slot in pied-piped finite RCs does not seem to be more productive in more advanced varieties.

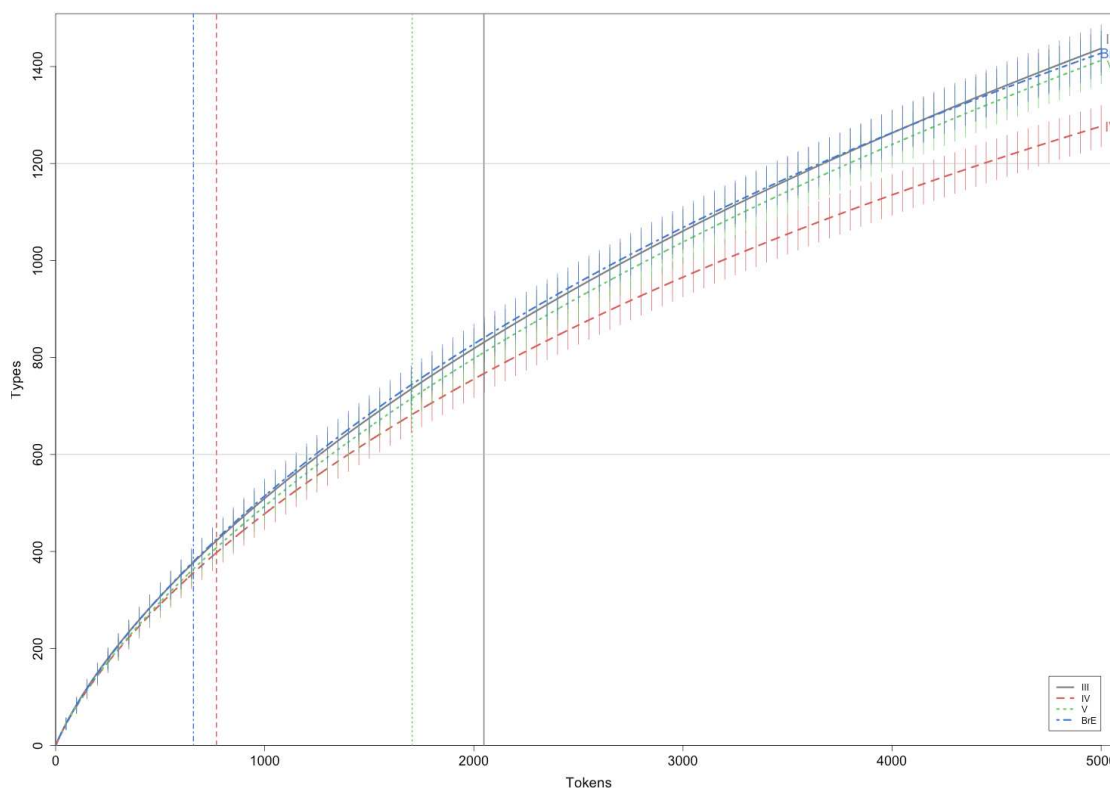


Figure 5.31: The productivity of the ANTECEDENT slot in pied-piped finite RCs (LNRE models)

One reason for this effect could be that V + P combinations (e.g. *stand for*, *deal with*) and FILLER + P combinations (e.g. *what...like*) frequently have idiomatic meanings. As PCEs tend to avoid constructions with non-compositional meanings (cf. 2.4.2), these patterns might be less deeply entrenched in less advanced varieties. Consequently, speakers of earlier Dynamic Model phases may rely on a smaller set of V + P chunks and FILLER + P combinations than speakers of varieties at higher stages. In contrast, the ANTECEDENT slot in pied-piped finite RCs is typically easy to interpret, which may lead to a high slot productivity in all varieties. Such an explanation was also suggested by Brunner (2021: 26), who attributed the results of his LNRE analysis of *into*-causatives, which turned out to be equally productive across all variety stages, to the fact that *into*-causatives can be interpreted fairly straightforwardly. In contrast, for example, the *way*-construction (Brunner & Hoffmann 2020) or the [*V the N_{taboo-word} out of*]-construction (Hoffmann 2020a), for which the Dynamic Model Productivity hypothesis was confirmed (cf. 2.2.3.2), are highly idiomatic. Moreover, the logistic regression analysis showed that finite RCs strongly favor pied-piping across all Dynamic Model

phases (cf. 5.2.4.2), which indicates that speakers of all varieties are likely to develop abstract templates for a pied-piped finite WH-relative clause construction. Also note that the curves in figure 5.31 may again be skewed by register effects. For example, the lower productivity of the ANTECEDENT slot in Phase IV may simply be due to the fact that the Phase-IV tokens might include more finite RCs from spontaneous contexts, in which speakers are more likely to rely on entrenched chunks (cf. 5.3.1). However, as the low token numbers of informal finite RCs did not permit an LNRE analysis based on informal finite RCs only, potential register differences between variety stages could not be taken into account in the present models.

Table 5.61: χ^2 - and p -values of the LNRE models of the P + FILLER slot in pied-piped finite RCs

VARIETY(_TYPE)	χ^2	p
III	5.0566347928231	0.167696730942035
IV	9.05675472104306	0.028545741448994
V	5.99037962800749	0.11207921634493
BrE	5.32838032364336	0.149271399434995

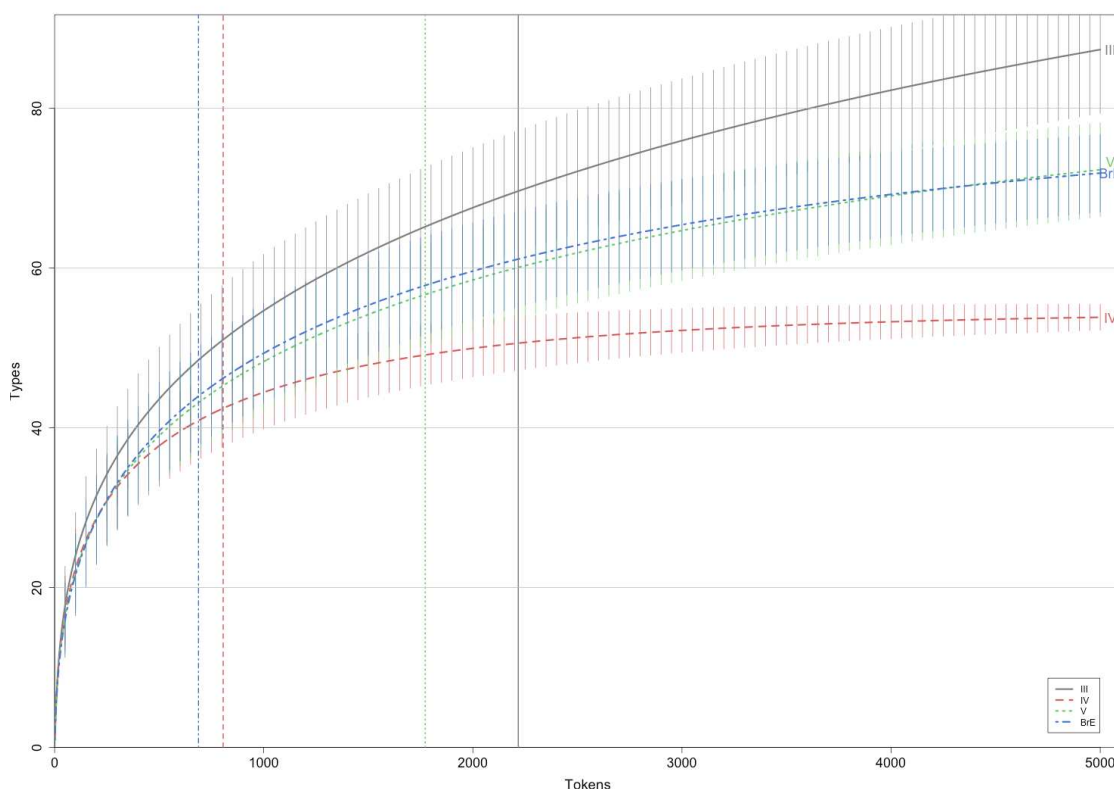


Figure 5.32: The productivity of the P + FILLER slot in pied-piped finite RCs (LNRE models)

The second LNRE analysis of pied-piped finite RCs focused on the combined P + FILLER slot. Note, however, that the results of the chi-squared test indicate that the goodness of fit of the Phase-IV model is not optimal (as $p < 0.05$; cf. table 5.61), which suggests that the results for Stage IV should not be taken too seriously.³⁴ Moreover, the relatively large confidence intervals in figure 5.32 indicate that the results for all variety phases should be taken with caution. As the number of possible P + WH combinations in finite RCs is much more restricted than e.g. the number of ANTECEDENT types, the occurrence of hapax P + WH combinations may have a very strong impact on the curves of the individual variety types. However, even when disregarding the curve of Phase IV, once again, the results are not in line with the Dynamic Model Productivity hypothesis. Instead, Phase III emerged as the most productive variety type, while the productivity of Phase V and BrE is considerably lower. Note that this effect cannot be attributed to sample size as BrE (688) has much lower token numbers than Phase V (2,217; cf. table A.29 in the Appendix).

Although this result may seem surprising at first, it is fully compatible with the assumption that less advanced varieties exhibit more variation in the preposition slot of pied-piped finite WH-relative clause constructions since meso-constructions that are deeply entrenched in the Stage-V varieties, such as a pied-piped *in-which* relative clause construction, are less deeply entrenched in the mental grammars of speakers of earlier Dynamic Model phases, who are exposed to less input (cf. 5.3.3). The fact that prepositions are frequently used in a non-literal sense and that they carry less meaning than e.g. antecedent nouns may make them prone to innovation. Furthermore, the high productivity of the P + WH slot in Phase III may reflect a processing principle. Investigating whether the presence of variable contexts is correlated with production difficulties, Gardner et al. (2021: 28) found that, “in certain contexts, the availability of multiple variants actually facilitates production”. Just like “hitting any target increases when the number of targets on the firing range is increased” (Gardner et al. 2021: 28), the availability of multiple P + WH patterns from which speakers can choose may thus facilitate processing. Consequently, the high productivity of the P + WH slot in Phase III might also be due to the fact that variation can lead to a reduction in processing difficulty.

³⁴ Using different cost functions and minimization algorithms did not improve the goodness of fit.

5.4.4 Summary

In sum, hypothesis V, which assumed that the productivity of preposition placement constructions is correlated with the evolutionary stage of a variety, was confirmed in parts. With respect to preposition stranding, the LNRE models provide support for the Dynamic Model Productivity hypothesis:

The LNRE models analyzing the productivity of the V + P slot (cf. 5.4.1) and the FILLER + P slot in interrogatives and free RCs (cf. 5.4.2) showed that these stranding constructions are most productive in the Phase-V and least productive in the Phase-III varieties, with Phase IV occupying an intermediate position. Phase-V speakers can thus be said to have more abstract representations of stranding constructions than speakers of less advanced varieties.

At the same time, hypothesis V was not confirmed with respect to pied-piped finite WH-relative clause constructions, i.e. a prototypical pied-piping context:

Regarding the productivity of the antecedent slot, the LNRE models did not identify clear differences between higher and lower Dynamic Model stages.

The least advanced varieties exhibit the highest productivity of the P + FILLER slot in finite RCs, which may be due to the weaker entrenchment of constructions in earlier Dynamic Model phases.

So far, the study has explored the use of preposition stranding and pied-piping constructions across varieties and variety phases. The final section of this corpus-based analysis, which aims to shed more light on qualitative constructional innovations in New Englishes, will focus on non-standard types of preposition placement.

5.5 Qualitative constructional innovations

In order to investigate whether innovative uses of preposition placement constructions are motivated by processing factors, second-language learning principles or cross-linguistic influence, all non-standard tokens of Phases III and IV, which make up the majority of all instances of non-standard preposition placement in the dataset (544, 91.12%; cf. table 5.63), were examined with the help of an HCFA which tested whether certain types of non-standard preposition placement are significantly associated with certain clause types, PP types, text types or varieties. Note that these innovations were not analyzed with the help of a logistic regression model because the low token numbers would have made it impossible to investigate effects of individual varieties or PP types. Furthermore, some factor levels had to be recoded and certain tokens had to be eliminated.

Due to the low frequencies of mismatching and doubling tokens (cf. table 5.63), these tokens were combined into a single level (“doubling”). The 75 cases with extra prepositions had to be excluded from the HCFA as the factor PP type could not be coded for these tokens. However, all extra cases will be examined in a separate analysis at the end of this section. Likewise, the 8 “other” tokens from HKE, IndE, PhilE, SgE and KenE, which all involve postpositions, had to be removed. While one of these tokens consists of the WH + P combination *which in* (5.60a), most of these cases involve the strings *where from* (3; cf. 5.60b-c) and *what for* (4; cf. 5.60d-e). Interestingly, 4 of these tokens with postpositions come from IndE. As both the Dravidian and the Indo-Aryan languages spoken in India have postpositions, this effect might be due to cross-linguistic influence. However, the low token numbers did not allow a closer investigation of this phenomenon.

- (5.60) (a) *he directed the uniquely Chinese collaboration entirely on the mainland on the relatively tight budget and in a language **which in** the novel and script were written* <ICE-HK:S2B-033#37:1:A>
- (b) ***Where from** did you bring it* <,> ? <ICE-IND:S1A-075#255:1:A>
- (c) ***Where from** in the world and when the HIV first emerged is not known* <ICE-EA-T:W2A021T>
- (d) *Ya what it's my if my if my aunty if my sister they all don't go **what for** I go hor* <ICE-SIN:S1A-007#101:1:A>
- (e) *You know <,> till the very end <,> Denis doesn't understand that **what for** he is being sent to prison* <ICE-IND:S1B-017#180:1:A>

Furthermore, the factor CATEGORY was recoded into the two levels “less spontaneous” and “more spontaneous” as it was expected that qualitative innovations

should be most likely to surface in more spontaneous usage contexts in which speakers do not have time for metalinguistic reflection. All ICE categories in which speakers' linguistic choices are influenced by time pressure during production, namely private dialogue, public dialogue, unscripted monologue and written as spoken, were categorized as "more spontaneous", while the written texts as well as the spoken categories broadcast news and scripted monologue were assigned to the level "less spontaneous" (cf. table 5.62).

Table 5.62: Recoding of CATEGORY into "less/more spontaneous"

	Less spontaneous	More spontaneous
Spoken	Broadcast news Scripted monologue	Private dialogue Public dialogue Unscripted monologue
Written	Private correspondence Business correspondence Non-professional writing Academic writing Non-academic writing Reportage Instructional writing Persuasive writing Creative writing	
ICE-EA only	Legal presentations	Written as spoken

The results of the HCFA are given in table 5.64. Note that only the most relevant significant configurations will be discussed here, while a full overview of all significant types and antitypes is provided in table A.30 in the Appendix. The first two significant types show that non-standard tokens occur more frequently than expected by chance in HKE and IndE. Furthermore, non-standard types of preposition placement are significantly associated with more spontaneous contexts, in which processing constraints can be assumed to have the strongest effect. That finite RCs emerged as a significant type can be attributed to the complexity of relative clause constructions. Finite RCs being the most complex type of filler-gap construction investigated in the present study (cf. 3.2.1), it is not surprising that these constructions are particularly likely to be affected by qualitative innovations. The PP types which are significantly associated with non-standard preposition placement are optional complements, prepositional verbs and respect PPs. As L2 Englishes have a preference for transparent form-meaning mappings, these PP types might be particularly prone to innovation since optional complements and prepositional verbs frequently have idiomatic meanings, and prepositions in respect PPs

Table 5.63: Non-standard preposition placement types across *VARIETY* (absolute and relative frequencies)

VARIETY (VARIETY TYPE)	Doubling		Mismatching		Missing		Resumptive		Unexpected		Extra		Other		Total	
CanE (V)	1	10.00%	0	0.00%	6	60.00%	0	0.00%	1	10.00%	2	20.00%	0	0.00%	10	100%
BrE (V)	1	14.29%	1	14.29%	3	42.86%	0	0.00%	0	0.00%	2	28.57%	0	0.00%	7	100%
IrE (V)	4	17.39%	1	4.35%	7	30.43%	2	8.70%	3	13.04%	6	26.09%	0	0.00%	23	100%
NZE (V)	3	23.08%	1	7.69%	4	30.77%	1	7.69%	2	15.38%	2	15.38%	0	0.00%	13	100%
JamE (IV)	1	2.50%	1	2.50%	18	45.00%	6	15.00%	11	27.50%	3	7.50%	0	0.00%	40	100%
SgE (IV)	11	13.92%	3	3.80%	42	53.16%	4	5.06%	11	13.92%	7	8.86%	1	1.27%	79	100%
KenE (III)	11	23.40%	2	4.26%	18	38.30%	1	2.13%	10	21.28%	5	10.64%	0	0.00%	47	100%
TanE (III)	3	4.29%	3	4.29%	31	44.29%	1	1.43%	16	22.86%	15	21.43%	1	1.43%	70	100%
HKE (III)	2	2.30%	1	1.15%	57	65.52%	6	6.90%	16	18.39%	4	4.60%	1	1.15%	87	100%
IndE (III)	8	8.70%	4	4.35%	48	52.17%	2	2.17%	18	19.57%	8	8.70%	4	4.35%	92	100%
NigE (III)	4	5.26%	0	0.00%	31	40.79%	8	10.53%	20	26.32%	13	17.11%	0	0.00%	76	100%
PhilE (III)	5	9.43%	3	5.66%	15	28.30%	0	0.00%	21	39.62%	8	15.09%	1	1.89%	53	100%
Column total	54	9.05%	20	3.35%	280	46.90%	31	5.19%	129	21.61%	75	12.56%	8	1.34%	597	100%

Table 5.64: HCFA results for *VARIETY * NONSTANDARD_TYPE * CLAUSE_TYPE * PP_TYPE * CATEGORY* (only the most relevant significant configurations are shown)

VARIETY	NONSTANDARD_TYPE	CLAUSE_TYPE	PP_TYPE	CATEGORY	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
HKE	.	.	.	.	82	58.875	9.0831	>	0.008859014494957	**	0.056
IndE	.	.	.	.	79	58.875	6.8792	>	0.02512640814245	*	0.049
.	.	.	.	more_spontaneous	309	235.5	22.9395	>	1.18587388374951E-11	***	0.312
.	.	finite RC	.	.	296	94.2	432.3062	>	2.83486495494475E-90	***	0.536
.	.	.	optional complement	.	196	26.1667	1102.2922	>	7.90879658118903E-115	***	0.382
.	.	.	prepositional X	.	82	26.1667	119.1345	>	1.08571582840914E-18	***	0.126
.	.	.	respect	.	60	26.1667	43.7461	>	4.25382329009493E-08	***	0.076
.	missing	.	.	.	260	117.75	171.8477	>	1.21337031702316E-43	***	0.403
.	resumptive	.	.	.	27	117.75	69.9411	<	4.36341003909697E-28	***	0.257
.	doubling	.	.	.	61	117.75	27.3508	<	1.5330270557E-10	***	0.161
.	missing	free RC	optional complement	more_spontaneous	30	9.9465	40.4306	>	0.000105210167782	***	0.043
IndE	missing	main clause interrogative	.	more_spontaneous	17	4.738	31.7342	>	0.002784723831293	**	0.026
.	unexpected	finite RC	.	less_spontaneous	51	26.587	22.4168	>	0.000339306162274	***	0.055
.	doubling	finite RC	prepositional X	less_spontaneous	14	2.2956	59.6763	>	9.6451080907508E-05	***	0.025

are typically used in an abstract sense (cf. 5.3.3). Furthermore, out of all non-standard types of preposition placement, it is the missing tokens which occur more frequently than expected by chance. This effect, which can be interpreted as a typical second-language learning simplification process (cf. 2.4.2), confirms previous studies on preposition placement which also found that L2 learners tend to omit prepositions in filler-gap constructions (see e.g. Kao 2001: 203; Sadighi, Parhizgar & Saadat 2004: 22, 25; Suárez-Gómez 2014: 260; cf. 3.3.3). However, considering that some local languages such as the Chinese languages spoken in Hong Kong or Singapore have relative clause constructions in which prepositions are omitted (cf. 3.4.2), at least in parts, this effect may also be due to cross-linguistic influence.

Table 5.65: HCFA results for VARIETY/VARIETY_TYPE * V of missing tokens with lexicalized/complement_like PPs (only significant results are shown)

VARIETY	VARIETY_TYPE	V	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
.	.	talk	16	1.6239	127.2691	>	1.50418748546E-09	***	0.082

In addition, the HCFA shows that tokens with missing prepositions are particularly associated with free RCs, optional complements and more spontaneous usage situations. As this indicates that prepositions are frequently omitted in constructions that involve entrenched V + P chunks, all missing tokens with lexicalized or complement_like PPs (cf. table 5.9) were subjected to an additional HCFA which tested whether certain verb-preposition strings are particularly likely to be affected by simplification in a certain variety or variety type. The only significant configuration that emerged as significant suggests that, across all PCEs, the verb *talk* appears without a preposition more frequently than expected by chance (cf. table 5.65). As shown in (5.61), *talk* is frequently used transitively in PCEs in cases in which Standard varieties would use *talk about*. Since *about* does not carry a lot of meaning, it might be considered redundant and thus be left out altogether in New Englishes.

- (5.61) (a) *What content did you **talk*** <ICE-HK:S1A-027#349:1:A>
 (b) *Well I don't I don't know what Rama has **talked** to you* <ICE-IND:S1A-096#86:1:B>

- (c) *That's what I was **talking*** <ICE-EA-T:S1B023T>
- (d) *Yes what do you want to **talk*** <ICE-SIN:S1A-006#4:1:B>
- (e) *that was the point that which we thought we would **talk** to you again* <ICE-JA:S1B-034#120:1:C>

Returning to the HCFA results given in table 5.64, another significant configuration involving missing tokens suggests that IndE tends to omit prepositions in main clause interrogatives. Since both the Indian-Aryan and the Dravidian languages spoken in India have postpositions, the use of prepositions may not be deeply entrenched in the mental grammars of speakers of IndE. As a result of cross-linguistic influence, these speakers might thus be likely to omit prepositions. Moreover, the examples in (5.62) and (5.63) indicate that prepositions are frequently dropped in stranded interrogative constructions involving V + P or FILLER + P combinations with idiomatic meanings (e.g. *apply to* (5.62a), *what...for* (5.62b)) and in pied-piped interrogative constructions in which Standard Englishes use prepositions to express abstract relationships (e.g. *in what way* (5.63a) or *in what manner* (5.63b)). As these constructions lack one-to-one form-meaning mappings, prepositions may have a high chance of being omitted.

- (5.62) (a) ***Which college** you will **apply*** <ICE-IND:S1A-051#42:1:A>
- (b) ***What** do you need it* <ICE-IND:S1A-056#97:1:A>
- (5.63) (a) ***What way** you found it different* <ICE-IND:S1A-042#160:1:A>
- (b) ***what manner** you people are approaching to the higher authorities?* <ICE-IND:S1A-083#69:1:B>

Besides this, the HCFA shows that unexpected tokens are significantly associated with finite RCs and less spontaneous contexts. The interaction of finite RCs and less spontaneous categories can be explained by the fact that the majority of the finite RCs in the dataset occur in less spontaneous, more formal text types. Furthermore, considering the fact that prepositions in finite RCs are typically pied-piped, the significant association between unexpected prepositions and finite RCs is perfectly compatible with the results of the LNRE models which suggested that the P + WH slot in pied-piped finite RCs is more productive in earlier Dynamic Model phases (cf. 5.4.3). While Stage-V varieties tend to rely on the string *in which* (cf. 5.3.3), speakers of PCEs use P + WH combinations more creatively. This effect is illustrated by the examples in (5.64), in which the prepositions *of* (5.64a), *on* (5.64b) and *to* (5.64c) would not be acceptable in Phase V. However, the use of the preposition *in* would render all these unexpected tokens

grammatical in Standard varieties (i.e. *an action programme in which* (5.64a), *areas in which* (5.64b), *a subject in which* (5.64c)).

- (5.64) (a) *Briefly I could say that we we have an action programme **of which** we are going to involve all uh NGOs in the country* <ICE-EA-T:S1B032T>
- (b) *I think <,> there are natural <,> areas **on which** co-operation will take place* <ICE-IND:S1B-039#76:1:B>
- (c) *Literature capital letter is a subject **to which** you study the writings and styles of the writers* <ICE-HK:S1B-019#35:1:A>

The HCFA of the non-standard tokens thus provides support for the claim that PCEs exhibit more variation in the preposition slot of pied-piped WH-relative clause constructions due to the weaker entrenchment of constructions (cf. 5.3.3). As suggested by Hoffmann's (2021) constructionist explanation of structural nativization in New Englishes (cf. 2.2.3.1), these qualitative innovations affect the meso-constructional level, i.e. the preposition slot of partly schematic ANTECEDENT + P + WH constructions (such as an *area-P-which* construction). Importantly, though, given their low frequencies, innovations such as (5.64a-c) should again be treated as constructs that are not conventionalized in the respective varieties.

Now, even though the HCFA in 5.3.3 indicated that e.g. the high-frequency string *in which* is less deeply entrenched in less advanced varieties, some of the unexpected tokens in the PCE data may also be explained by a preference for frequent forms. As shown in table 5.66, almost two thirds of the unexpected tokens involve the prepositions *on*, *in*, *of* and *to*. Figure 5.33, which visualizes the relative frequencies of the five most frequent prepositions per variety, shows that these prepositions are high-frequency prepositions in all ICE components (for the absolute preposition frequencies, see table A.31 in the Appendix). Recall that Schmied (2004: 934) found that East African Englishes tend to overuse *in* in cases in which Standard Englishes would choose more specific prepositions (cf. 2.4.2). In a similar vein, in (5.65), for example, the use of the string *of which* instead of the standard variant *at which* may be attributed to the high frequency of *of*, which is by far the most frequent preposition in ICE-EA-T.

- (5.65) *As you might be aware Honourable Minister a lot has taken place since Arusha National Conference in August nineteen ninety-four **of which** we had a privilege of having you as our guest of honour too* <ICE-EA-T:S2B030T>

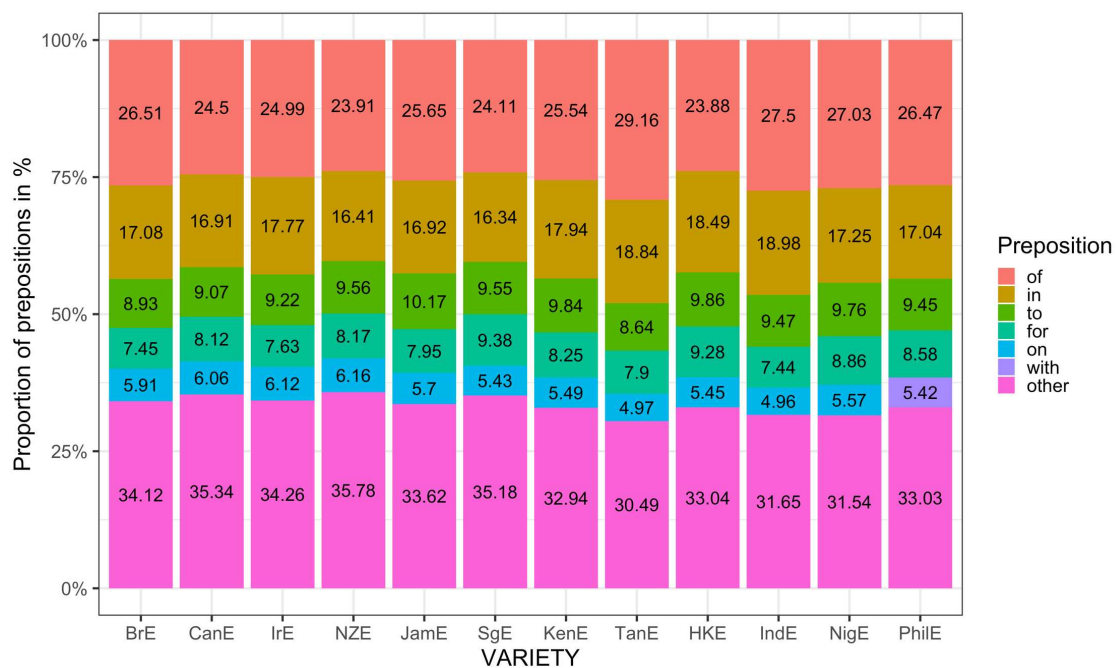


Figure 5.33: Relative preposition frequencies per ICE variety

Table 5.66: Prepositions in unexpected tokens (absolute and relative frequencies)

PREPOSITION	Abs. freq.	Rel. freq.
on	26	20.16%
in	21	16.28%
of	20	15.50%
to	15	11.63%
with	8	6.20%
for	7	5.43%
at	6	4.65%
by	6	4.65%
from	5	3.88%
into	3	2.33%
over	3	2.33%
out of	2	1.55%
under	2	1.55%
about	1	0.78%
around	1	0.78%
out	1	0.78%
upon	1	0.78%
within	1	0.78%
Total	129	100.00%

Moreover, some of the stranded unexpected tokens may be motivated by a preference for more explicit constructions. In (5.66a), for example, the preposition *into* may be used instead of *to* as the former preposition is more explicit. According to Quirk et al. (1985: 675), *to* typically expresses movement to a destination, whereas *into* emphasizes “the combination of destination + dimension”. Consequently, in (5.66a) the preposition *into* may be chosen in order to emphasize the goal of the writing process, namely getting through to the companies for which the person would like to work. Likewise, in (5.66b) *sit into* may be preferred over *sit in* to highlight the body movement that is required when a person puts on a harness.

- (5.66) (a) *I've got a friend uh I don't know how many companies she **wrote into** you know and then not one interview* <ICE-SIN:S1A-046#28:1:A>
- (b) *Then he has his uh harness which he **sits into** to which it 's attached a carabina and this device which is called figure-of-eight descender* <ICE-SIN:S2B-015#44:1:A>

With respect to the other types of non-standard preposition placement, the HCFA reveals that resumptive and doubling tokens occur less frequently than expected by chance. Even though the HCFA did not identify variety-specific effects regarding the use of resumptive pronouns, interestingly, PhilE is the only New English variety which does not have a single resumptive token. As the Philippine languages do not have preposition placement constructions with resumptive pronouns, this might be due to the fact that speakers of PhilE do not encounter resumptive structures in their L1 input, whereas resumptive pronouns are more frequent in the local languages of Singapore, Hong Kong, Nigeria, Kenya and Tanzania (cf. 3.4.2). At the same time, resumptive tokens surface in ICE-IND and ICE-JA even though the L1s spoken in Jamaica and India do not license preposition placement constructions with resumptive pronouns either. Resumptive pronouns being a common feature of L2 Englishes (cf. 3.3.3), it is thus impossible to decide whether the few resumptive tokens contained in the present dataset are motivated by cross-linguistic influence or second-language learning principles.

Although the HCFA identified doubling per se as an antitype, the association between preposition doubling, finite RCs, prepositional verbs and less spontaneous text types emerged as a significant type. While the association between finite RCs and less spontaneous text types can again be explained by the fact that most of the finite RC tokens occur in more formal, written categories, the association between doubling, finite RCs and prepositional verbs can be linked to a processing effect (Hoffmann 2011: 210, 225;

cf. 3.3.3). Remember that finite RCs strongly prefer pied-piping, while prepositional verbs are easier to process in stranded constructions in which verb-preposition string are not broken up. In cases such as (5.67), the use of both a stranded and a pied-piped preposition thus leads to a reduction in processing cost. While the first *from* reduces the risk of a garden path effect (as it avoids that *which* is identified as the direct object of *draw*), the second preposition ensures that the verb-peposition string *draw from* is maintained (also see Hoffmann 2011: 194).

(5.67) *Its contributions to English **from which** ESP draws **from** is the ‘‘slot and filter’’ form* <ICE-NIG:ex_23>

Furthermore, as suggested in 3.3.3, some of the doubling structures with prepositional verbs in the dataset can be attributed to the entrenchment of (ANTECEDENT +) P + WH combinations. For example, the string *method through which* is likely to be entrenched as a prototypical means/instrument adjunct PP chunk in speakers’ mental grammars although it does not function as a means/instrument PP in (5.68). Consequently, it has a high chance of being activated even in structures such as (5.68), in which *method* is the complement of a prepositional verb. At the same time, the string *go through* (meaning ‘experience’) can also be assumed to be stored as a complex lexical item. Example (5.68) can thus be considered the result of the simultaneous activation of two entrenched meso-constructions, viz. a pied-piped *method-through-which* and a stranded *go-through* construction. Also note that the doubling tokens involving the preposition *from* in (5.69a-c) provide further evidence for the entrenchment of a partly schematic pied-piped *from-where* construction in the less advanced varieties (cf. 5.3).

(5.68) *According to Kemery, there is one **method through which** all scientific theories **go through** in their development.* <ICE-SIN:W1A-014#62:2>

(5.69) (a) *I did not know **from where** the identifying witnesses had **come from*** <ICE-EA-K:S1BCE07K>

(b) *I say do you know **from where** I **came from*** <ICE-NIG:con_14:a273>

(c) ***From where** I **come from**, we eat grass* <ICE-SIN:W2F-015#141:1>

Finally, it should be noted that, even though the HCFA investigated non-standard preposition placement in PCEs, doubling tokens such as example (5.70), which includes the pied-piped chunk *case in which* and the stranded string *involved in*, also occur in the

most advanced varieties. However, as the low token numbers of Phase V (9 doubling, 3 mismatching; cf. table 5.63) did not allow for a statistical analysis, a thorough investigation of this phenomenon requires a larger database.

- (5.70) *It concerns a **case in which** British rider Harvey Smith was **involved in** in a dispute over alleged damage to his land by greenkeepers from a nearby golf course* <ICE-IRE:S2B-013\$G>

Table 5.67: *P + WH combinations in mismatching tokens of Phases III and IV (absolute and relative frequencies)*

P + WH	Abs. freq.	Rel. freq.
in_which	6	37.50%
of_which	2	12.50%
on_which	2	12.50%
about_which	1	6.25%
for_which	1	6.25%
in_what	1	6.25%
till_when	1	6.25%
to_which	1	6.25%
with_whom	1	6.25%
Total	16	100.00%

Since doubling and mismatching tokens had to be combined into a single level for the HCFA, next, a closer look was taken at the 16 mismatching tokens of Phases III and IV. Although these tokens could not be subjected to an additional statistical analysis, support for the assumption that mismatching structures frequently result from the activation of partly schematic stranding and pied-piping constructions (cf. 3.3.3) comes from the fact that all but two of these tokens contain optional (5.71a) or obligatory complement PPs (5.71b) or prepositional verbs (5.71c), which are likely to involve V + P chunks (e.g. *proud of* (5.71a), *rely on* (5.71c)).

- (5.71) (a) *And I think that is something **about which** every Indian <,,> should feel proud **of*** <ICE-IND:W2F-010#150:1>
- (b) *we look at the environment likewise **in which** the employee is **at*** <ICE-PHI:S2A-023#103:1:A>
- (c) *hence failure to run the industries due to the lack of money to input oil **on which** it heavily rely **at*** <ICE-EA-T:W1A018T>

Moreover, as shown in table 5.67, more than a third of the mismatching tokens contain the high-frequency string *in which*. Even though no final conclusions can be drawn from the low frequencies, this observation is consistent with Riley & Parker's (1986: 302) finding that structures with mismatched prepositions frequently involve the high-frequency preposition *in* (also see Jach 2021: 366).

Table 5.68: *P + WH combinations in extra tokens (absolute and relative frequencies)*

P + WH	Abs. freq.	Rel. freq.
of_which	20	36.36%
in_which	15	27.27%
on_which	4	7.27%
with_which	3	5.45%
by_which	2	3.64%
through_which	2	3.64%
at_which	1	1.82%
for_which	1	1.82%
for_whom	1	1.82%
into_which	1	1.82%
to_what	1	1.82%
to_which	1	1.82%
to_whom	1	1.82%
upon_which	1	1.82%
with_whom	1	1.82%
Total	55	100%

Turning to the extra tokens, which could not be included in the HCFA, the majority (70.13%, 54) of the structures in which no preposition would be used in Standard Englishes involve pied-piped prepositions (cf. table A.32 in the Appendix). While the low token numbers only permit a qualitative analysis, it is worth noting that almost two thirds (63.63%, 35) of these pied-piping cases include the high frequency strings *of which* (5.72) and *in which* (5.73) (cf. table 5.68). Just like the mismatching tokens, these structures thus seem to result from the activation of entrenched P + WH chunks.

- (5.72) (a) *there is uh a proverb **of which** says that repetition is the mother of knowledge* <ICE-EA-T:S1B049T>
- (b) *You see according to the hotel record **of which** you also confirmed you left Conrad hotel on the ninth of <,> November nineteen ninety two* <ICE-HK:S1B-068#33:1:A>

- (5.73) (a) *I shall look at the some of the factors or forms **in which** I feel are important uh for motivation for learning* <ICE-SIN:S2A-050#105:1:A>
- (b) *The step **in which** Tanzania government has already undertook in mgololo paper industries* <ICE-EA-T:W1A018T>

Furthermore, some of the pied-piped extra cases may be due to the entrenchment of partly schematic ANTECEDENT + P + WH constructions such as a *person-to-whom* (5.74) or a *point-at-which* (5.75) construction:

- (5.74) *If financial position is poor then they must go to work but they leave their small baby in hands of some <}> responsible </}> **person to whom** you trusts fully* <ICE-IND:W1A-001#118:2>
- (5.75) *And this has really been one of **the points** <,> **at which** many of the professional <,> community builders have had with the political directorate* <ICE-JA:S2A-001#31:1:A>

The stranded extra tokens include the innovations *discuss about* (3x; (5.76)) and *request for* (1x; (5.77)), which have been found in many New Englishes and which can be said to result from a process of remapping that leads to more transparent form-function mappings (cf. 2.4.2):

- (5.76) *therefore I paused for a for some time and tried to listen what they were **discussing about*** <ICE-EA-K:S1A008K>
- (5.77) *that was what they **requested for*** <ICE-NIG:bdis_17:a1191>

Moreover, examples such as *target at* (instead of *target* in (5.78)) and *follow upon* (instead of *follow* in (5.79)) may reflect a preference for more explicit constructions in PCEs (see 2.4.2). In both (5.78) and (5.79), the prepositions explicitly express the movement implied by the verbs (see Nesselhauf 2009: 20-21). Note that *follow upon* in (5.79) is additionally followed by a resumptive pronoun. Furthermore, analogy may play role in the emergence of *belong to* in (5.80). Since (5.80) has the meaning ‘a place where you feel comfortable’, Standard Englishes would use the variant *Where you belong*. However, as speakers of PCEs frequently encounter the prepositional verb *belong to* with the meaning ‘be owned by’ in constructions such as *Who does this book belong to?*, they may also start to use *belong to* with the meaning ‘feel comfortable’.

- (5.78) *Whom are you **targeting at*** <ICE-EA-T:S1B046T>

(5.79) *you know every youth them are follow upon their DJs or them out them favourite artist or whatever you **follow upon** it like it's your life* <ICE-JA:S2A-035#64:1:A>

(5.80) *Where you **belong to*** <ICE-EA-T:S1B033T>

In sum, the results presented in this section confirmed hypothesis VI, which assumed that PCEs should be characterized by qualitative constructional innovations at the meso-constructional level that are motivated by processing factors and second-language learning principles. The fact that qualitative innovations in PCEs typically emerge in constructions with lexicalized verb-preposition strings and in constructions in which prepositions have very abstract meanings suggests that New Englishes have a preference for one-to-one form-function mappings. Moreover, the missing tokens make up the most frequent type of nonstandard preposition placement in the ICE data, which shows that PCEs are frequently characterized by simplification processes that are characteristic of second-language acquisition. While many structures with doubled or extra prepositions can be said to result from the activation of entrenched P + WH or V + P chunks, qualitative constructional innovation can frequently also be attributed to processing factors: Not only do the doubling and resumptive tokens indicate that New Englishes prefer structures that involve a reduction in processing effort, but also the HCFA demonstrated that the cognitively most demanding contexts, i.e. more spontaneous text types and relative clause constructions, are most likely to be affected by innovations. Finally, the unexpected and extra tokens showed that innovation typically takes place in the preposition slot of pied-piped ANTECEDENT + P + WH (e.g. *area P which*) or stranded V + P (e.g. *target at*) strings. The results thus support the prediction that qualitative innovations in PCEs tend to surface in partly schematic constructions.

5.6 Summary and discussion

This chapter presented a corpus-based analysis of preposition placement in World Englishes, which investigated the use and entrenchment of preposition stranding and pied-piping constructions at all levels of abstraction. The results are summarized in table 5.69.

Table 5.69: Summary of study results (Hypotheses I - VII)

Hypothesis	Result	Explanation
I	✓	III-V: pied-piping with finite WH-relatives, adjunct PPs, NP-contained PPs, passive constructions, long FGDs
	✓	III-V: stranding with lexicalized verb-preposition strings
II	✓	III: more pied-piping in adjunct interrogatives
III	✓	III: pied-piping regardless of formality
		IV-V: pied-piping in more formal, stranding in less formal/personal contexts
IV	✓	III: more V + P chunks in non-prototypical stranding contexts (stranded finite RCs)
	✓	III: fewer non-prototypical FILLER + P interrogative combinations
	X	III: more variation in the P slot of pied-piped finite RCs
	X	V: stronger reliance on <i>in which</i> , <i>way in which</i> in pied-piped finite RCs
V	✓	productivity of the V + P slot in stranded constructions
	✓	productivity of the FILLER + P slot in stranded free RCs and interrogatives
	X	productivity of the ANTECEDENT slot in pied-piped finite RCs
	X	productivity of the P + WH slot in pied-piped finite RCs
VI	✓	doubled, mismatching, missing, resumptive, unexpected, extra tokens
	✓	III: <i>how...look like</i>
VII	✓	III: pied-piping in interrogatives in HKE, IndE, PhilE, KenE, TanE
	✓	stranding in finite RCs in SgE
	✓	pied-piping in IndE

Hypothesis I was confirmed as the logistic regression models (cf. 5.2) suggested that speakers of all variety types prefer constructions which are easy to process, namely the stranded free RC construction, the stranded WH-interrogative construction, and the pied-piped finite WH-relative clause construction. Moreover, pied-piping is preferred with NP-contained PPs, with long FGDs, with the passive, and with respect, manner and degree PPs. In contrast, stranding is only favored in constructions in which the verb and the preposition are closely associated with each other. At the same time, the results indicate that less advanced varieties generally have a stronger preference for pied-piped constructions than speakers of more advanced varieties. Not only did the logistic regression analysis demonstrate that lower Dynamic Model stages have a weaker stranding preference with adjunct PPs and with free RCs and interrogatives than varieties at higher stages, but also the HCFA revealed that only Phase III has entrenched FILLER + P chunks for pied-piped adjunct interrogatives (cf. 5.3.2). At the same time, in line with hypothesis IV, which predicted that less advanced varieties should favor prototypical constructional fillers, only Phase V has partly schematic FILLER + P constructions for stranded adjunct interrogatives, which should be dispreferred from a processing perspective. As this implies that speakers of Stage-III varieties encounter adjunct WH-interrogative constructions primarily with pied-piped prepositions, it is plausible to assume that the Phase-III (and possibly also the Phase-IV) varieties possess a pied-piped adjunct WH-interrogative construction that can be attributed to processing factors (cf. 5.3.2). Speakers of IndE, who have an almost equal preference for stranding and pied-piping in interrogatives (cf. 5.2.3.2), can be said to have the most deeply entrenched pied-piped WH-interrogative construction. The results thus also confirmed hypothesis II. The stronger preference for pied-piped adjunct interrogative constructions in Phase III suggests that speakers of varieties at earlier stages of evolution, who are exposed to less input and for whom the use of English is associated with higher cognitive cost, have a stronger preference for constructions that involve a reduction in processing complexity.

Importantly, however, this stronger pied-piping preference with adjunct PPs in the less advanced varieties only turned out to be significant in the logistic regression analysis which included all clause types (cf. 5.2.1 and 5.2.2) but not in the one focusing on finite RCs (cf. 5.2.3). Moreover, apart from a very weak interaction effect between RESTRICTIVENESS and VARIETY_TYPE (cf. 5.2.3.2), the interaction of PP_TYPE and VARIETY_TYPE was the only interaction between a processing-related variable and VARIETY_TYPE that emerged as significant in the regression models. This suggests that processing-related differences between variety phases only emerge in environments

in which the processing load is not maximal, namely in WH-interrogative constructions with prototypical adjunct PPs such as time or location PPs. At the same time, in accordance with Hawkins' (2004: 3) PGCH, speakers of all variety phases strongly favor pied-piping in structures that are very difficult to process, i.e. in finite RCs, in constructions with NP-contained PPs and in structures with very long FGDs. Note, though, that the absence of more significant interaction effects between VARIETY_TYPE and processing-related variables may also be due to the fact that not more three-way interactions could be tested in the regression analysis. For instance, optional complement PPs contained in NPs (such as *the building site which this building has been **part of*** <ICE-IRE:S1B-012\$A>) may favor stranding more than NP-embedded PPs in which the noun and the preposition are not closely associated with each other. However, as the three-way interaction between PHRASE_TYPE, PP_TYPE and VARIETY_TYPE was too complex to be included in the regression models, the results regarding PHRASE_TYPE may be skewed by an unequal distribution of the PP_TYPE levels across the NP-tokens of the three variety phases. Similar effects may, of course, also apply to other variables.

At the same time, these results confirm the predictions made in 2.4.1.1 that processing-related differences between varieties are most likely to surface with respect to constructions whose competition is strongly driven by parsing-related variables. Recall that pied-piping with adjunct PPs reduces the risk of garden paths as it complies with Hawkins' (1999: 277) Avoid Competing Subcategorizers principle. In other words, pied-piped adjunct PP constructions increase parsing efficiency as they avoid misassignments so that the filler can be matched with the gap as early as possible. That processing-related differences between L1 and L2 varieties have been found in the present study and with respect to *that*-complementation (Wulff, Lester & Martinez-Garcia 2014; Wulff, Gries & Lester 2018) but not in studies on the dative and the genitive alternation (Bernaisch, Gries & Mukherjee 2014; Heller, Szmrecsanyi & Grafmiller 2017; Röthlisberger, Grafmiller & Szmrecsanyi 2017) may thus be due to the fact that, compared to preposition placement or *that*-complementation, the dative and the genitive alternation are not as strongly influenced by parsing-related variables. Moreover, disregarding pragmatic aspects, preposition stranding and pied-piping constructions are semantically equivalent. In contrast, the two dative and the two genitive constructions are not always semantically interchangeable, which may lead to similar constructional choices across all varieties. The results of the present study thus support the assumption that processing-related differences between L1 and L2 speakers are most likely to emerge in studies that analyze

competing constructions which are semantically interchangeable and which are strongly affected by variables that have an impact on parsing efficiency (cf. 2.4.1.1). Furthermore, it is important to remember that the present study investigated preposition placement in filler-gap constructions, i.e. a very complex construction type that makes very high demands on working memory. As outlined in 2.4.1.2, psycholinguistic research suggested a number of factors that may explain why these long-distance dependencies are more difficult to process for L2 speakers than for L1 speakers. Another reason that might account for the fact that the present study found that less advanced varieties have a stronger preference for prototypical constructions, whereas no such effect could be confirmed in studies on the dative and the genitive alternative could thus be the great complexity of filler-gap constructions.

Apart from processing factors, formality and topic also turned out to affect the competition between preposition stranding and pied-piping. Crucially, though, confirming hypothesis III, the logistic regression analysis revealed that formality has a stronger effect on varieties at higher Dynamic Model stages. While speakers of all variety phases can be said to store pragmatic associations between the pied-piped finite WH-relative clause construction and formal contexts, only Stages IV and V also have an abstract template for a stranded finite RC construction that is associated with informal contexts and more personal topics. In contrast, the pied-piping preference of the less advanced varieties is only slightly weaker in less formal contexts than in more formal ones (cf. 5.2.4.2). Moreover, in contrast to Hoffmann's (2011) corpus study, the present analysis did not only identify a formality effect for finite RCs but also for free RCs and interrogatives (cf. 5.2.2.2). The results thus generally confirm the assumption of major usage manuals such as Quirk et al.'s (1985) *Comprehensive grammar of the English language* that pied-piping is associated with formal contexts in both WH-relative and WH-interrogative constructions (cf. 3.1.1.1). However, as the formality effect for interrogatives is clearly weaker than for finite RCs (cf. figure 5.9), it is rather unlikely that speakers of the more advanced varieties also have a pied-piped interrogative construction that is associated with more formal contexts. Note that the absence of an informal stranded finite WH-relative clause construction in the less advanced varieties results can be explained by the interplay of input effects and processing factors: The formal input that is typical of Stage-III varieties can be expected to mostly contain constructions with pied-piped prepositions. Consequently, these speakers do not receive sufficient input for stranding constructions that would allow them to develop an abstract representation of a stranded finite WH-relative clause construction (cf. 5.2.4.2). At the

same time, speakers of less advanced varieties favor pied-piping due to processing reasons, which further contributes to the entrenchment of pied-piping constructions. However, as pointed out in 2.3, preemption only works when speakers possess constructional alternatives. As a result, the Phase-III varieties, which lack alternative finite WH-relative clause constructions from which they can choose, prefer pied-piping regardless of the context.

These effects are also summarized in figures 5.34 and 5.35, which give an overview of the abstract constructions that can be assumed to be entrenched in the construction networks of Phases III and V based on the results of the corpus study. Note that no figure is provided for the construction network of Phase IV since the low token numbers of the Stage-IV varieties made it impossible to draw definite conclusions about e.g. the entrenchment of a pied-piped WH-interrogative construction in these varieties. Just like in Hoffmann's (2011) account of the BrE and the KenE construction networks, the differences between Phases III and V can be explained by two factors: processing constraints and formality effects. While Phase III has a pied-piped adjunct interrogative construction which can be attributed to the stronger effect of processing constraints in the Stage-III varieties, Phase V has an informal stranded WH-relative clause construction that is associated with informal contexts. Importantly, this does not mean that Phase III never uses stranded finite WH-relatives or that pied-piped interrogatives are nonexistent in Phase V. Instead, as pointed out in 3.1.1.2, these structures can be produced by combining other constructions. In Phase V, pied-piped WH-interrogatives can be generated by the general WH-pied-piping construction which can also be used to produce pied-piped free relatives and exclamatives (cf. 3.1.1.1). At the same time, speakers of Stage-III varieties can use the general stranding construction to produce stranded finite WH-relatives. That the construction network of Phase V also includes a general stranding construction can be explained by the fact that stranded free RCs and stranded WH-interrogatives could, in theory, be generated by the same abstract construction. However, as pointed out in 3.1.1.2, from a usage-based perspective, it is more plausible to assume that speakers store additional lower-level constructions. Just like in section 3.1, the circle marking the general stranding construction in figures 5.34 and 5.34 thus indicates that this construction is slightly more abstract than the stranded free relative and the stranded WH-interrogative construction. The arrow connecting the stranding and the pied-piping families shows that all other constructions are interconnected by horizontal relations at the same level of abstraction (see 3.1). Note that, compared to the family of preposition placement construction presented in 3.1, these figures do not include representations for

clause types with categorical preposition placement preferences such as stranded *that*- or zero-relative clause constructions, which were not investigated in the present study. Likewise, topicalizations, which could not be extracted from the corpus data, and *it*-cleft constructions, whose token numbers were too low to be included in the statistical analysis, are excluded. Furthermore, the constructions shown in figures 5.34 and 5.35 can, of course, also be expected to be related to sentence neighbors such as other relative or interrogative constructions. However, since these relations were not investigated in the present analysis, they are not represented in figures 5.34 and 5.35.

Now, above, I have repeatedly argued that the results indicate that less advanced varieties generally have a stronger preference for pied-piped constructions. That preposition stranding, i.e. the variant that involves an increase in processing cost, is less deeply entrenched in varieties at earlier Dynamic Model stages was not only suggested by the logistic regression analysis but also by the LNRE models, which indicated that Stage V has the most abstract templates for stranding constructions (cf. 5.4.1, 5.4.2). Particularly in more informal, more spontaneous contexts, varieties at earlier stages rely more on entrenched verb-preposition strings than more advanced varieties. Besides this, the HCFAs demonstrated that Phase III uses more entrenched meso-constructions in stranded finite RCs (cf. 5.3.1). With respect to preposition stranding, the results thus confirmed hypotheses IV and V, which predicted that less advanced varieties should rely more on prototypical lower-level constructions, while Phase V should have the most abstract constructional representations. As pointed out above, the weaker entrenchment of preposition stranding in the less advanced variety can be attributed to the formal input that is typical of PCEs, and to the high processing complexity of stranding constructions. Since pied-piped constructions are easier to process, speakers of varieties at lower developmental stages only use stranded prepositions with lexicalized verb-preposition strings, which can be activated without effort (but not with e.g. stranded adjunct WH-interrogative constructions). However, due to the limited amount of input that is characteristic of lower Dynamic Model phases, speakers of less advanced varieties possess fewer and less deeply entrenched V + P chunks than speakers of varieties at higher stages. Sections 5.3.1 and 5.3.2 indicated that Stage V stores more V + *about*, V + *like* and V + *from* strings than Stage III, which, in contrast, shows a higher reliance on more substantive *what...talk about* and *how...look like* constructions. Put differently, in the less advanced varieties, there are fewer cases in which the use of stranded prepositions results in a processing advantage, which leads to an overall preference for pied-piping.

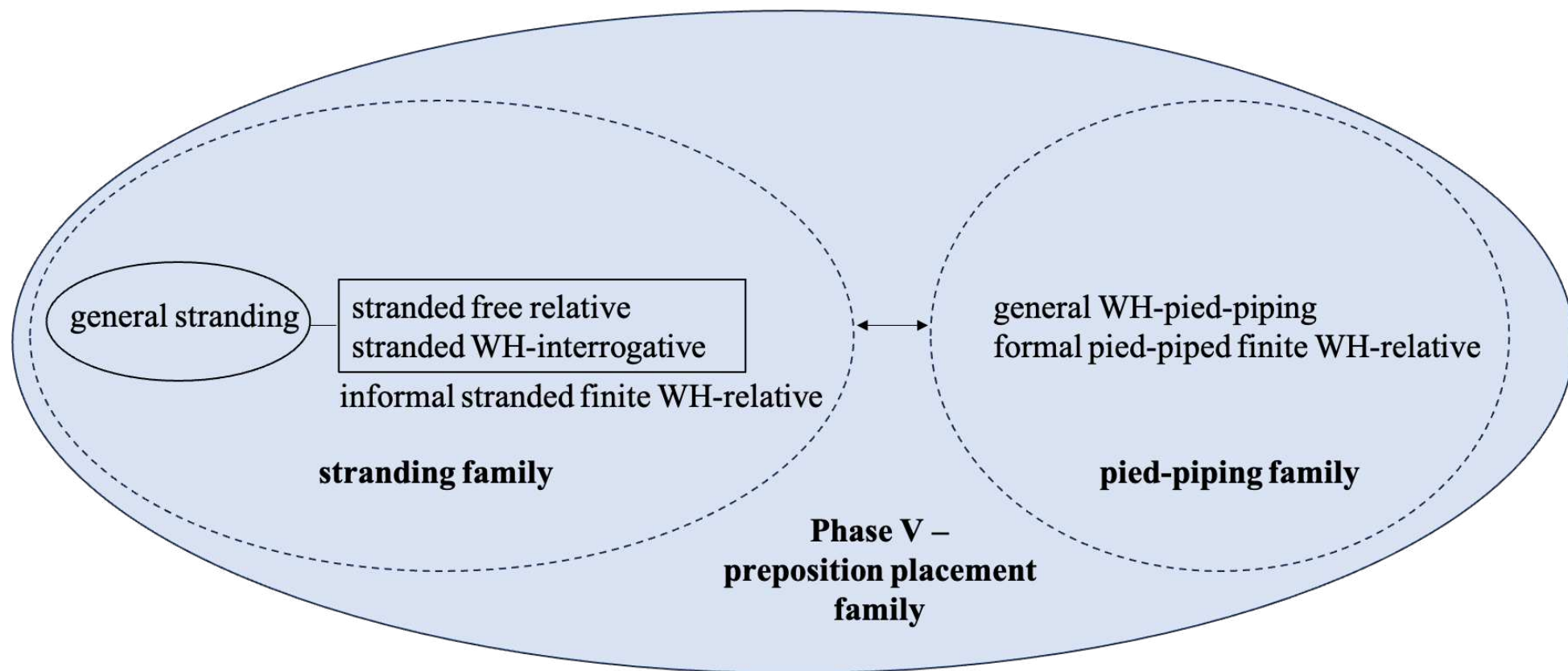


Figure 5.34: The preposition placement family of Phase V (variable clause contexts only)

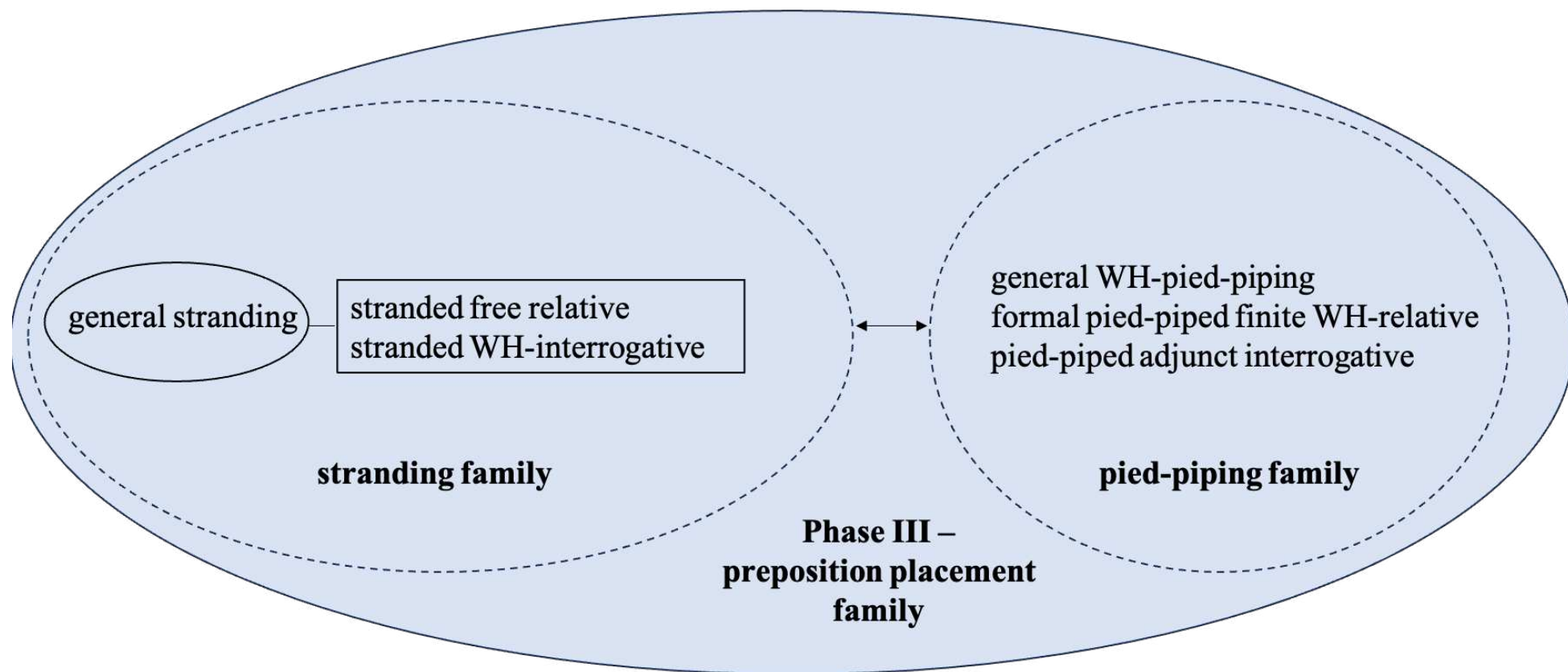


Figure 5.35: The preposition placement family of Phase III (variable clause contexts only)

While the LNRE analysis revealed that Phase V has more abstract stranding constructions than varieties at lower developmental stages, it did not identify significant differences between variety types with respect to the productivity of the antecedent slot in pied-piped finite RCs. Moreover, surprisingly, the combined P + WH slot turned out to be most productive in Phase III (cf. 5.4.3). This suggests that speakers of all variety types have abstract representations of pied-piped finite WH-relative clause constructions. That Hoffmann's (2021: 21) Dynamic Model productivity was not confirmed with respect to pied-piped finite RCs can be explained by several reasons. First of all, from a corpus-as-input view, the logistic regression analysis showed that speakers of all variety phases encounter lots of input for pied-piped finite RC constructions, which can be explained by the high processing complexity of relative clauses. Moreover, especially speakers of varieties at lower stages of evolution also receive input for P + WH combinations in other clause types such as pied-piped interrogatives. In addition, the HCFA's (cf. 5.3.3) and the analysis of non-standard tokens (cf. 5.5) showed that speakers of less advanced varieties exhibit more variation in the preposition slot of pied-piped finite WH-relative clause constructions, whereas Phase-V speakers strongly rely on entrenched meso-constructions such as a pied-piped *in-which* or *way-in-which* construction. While this effect is contrary to hypotheses IV and V, it is in accordance with Laporte's (2021: 321) finding that, "[a]t the intermediate level, [...] more advanced varieties exhibit less variation in the instantiations of constructions". According to Laporte (2021: 316), this effect can be explained by the fact that

less entrenched constructions do not necessarily entail a higher reliance on lower-level constructions. When considering realization patterns, that is, subtle formal variations, it is in fact less entrenched constructions that give rise to new patterns of instantiations. This is corroborated by the fact that [...] the majority of non-standard patterns that account for increased variation do not seem to be conventional patterns of the varieties and can therefore not be claimed to represent stored constructions that speakers of New Englishes draw on.

A similar suggestion was made by Wulff & Gries (2019: 902), who claimed that "the input learners receive is most varied and, hence, ambiguous or uncertain, opening more opportunities for nontargetlike choices". Since prepositions in pied-piped finite RCs frequently do not have concrete meanings, their use is less deeply entrenched in PCEs. Consequently, qualitative innovations in Phase III frequently affect the preposition slot of ANTECEDENT + P + WH patterns. Innovative uses of partly schematic constructions (e.g. *way of which* instead of the standard variant *way in which*) thus explain why the overall productivity of the P + WH slot is higher in the less advanced varieties. Just like in Laporte's (2021) study, the majority of the innovations identified in the present analysis

can be considered constructs of individual speakers that are not part of the construction networks of the respective varieties.

But why is it that only the P + WH slot but not the V + P slot turned out to be most productive in the least advanced varieties even though verb-preposition strings typically also involve prepositions that are not used in a literal sense? This might be due to the fact that, compared to P + WH combinations, V + P chunks tend to be more idiomatic. In the string *way in which*, for example, the preposition *in* does not carry a lot of meaning and the meaning of the construction does not depend on the preposition. Consequently, the meaning does not change if a speaker uses e.g. *way of which* or *way from which* instead of the standard variant *way in which*. In contrast, in stranding constructions with prepositional verbs such as *This is what the President stands for*, the V + P chunk has a non-compositional meaning (e.g. ‘represent’) that changes if the P slot is filled by a different preposition (e.g. *This is what the President stands on*). In a similar vein, Wulff & Gries (2019: 902) pointed out that “[i]diomatic strings are typically more formulaic, that is [...] constructional variability is low and predictability of what constitutes the right choice correspondingly is high“. While the present study found some innovative uses of verb-preposition combinations (cf. 5.5), the constant, idiomatic meaning of e.g. prepositional verbs may explain why the P slot of verb-preposition strings is not as prone to innovation in the Stage-III varieties as the P + WH slot of finite RCs. Furthermore, the lower productivity of the V + P slot in Phase III can be attributed to the fact that speakers of less advanced varieties have fewer entrenched idiomatic stranding constructions as PCEs tend to prefer transparent constructions with one-to-one form-meaning mappings (cf. 2.4.2). This can, for example, be seen in the fact that idiomatic *WH...like* and *WH...be up to* constructions turned out to be significantly associated with Phase V only (cf. 5.3.1, 5.3.2). While the slots of pied-piped finite RCs are equally productive across all Dynamic Model stages (with Phase III exhibiting the highest productivity of the P + WH slot), varieties at earlier stages rely more on a small set of entrenched verb-preposition strings in stranding constructions.

The claim that pied-piping is more deeply entrenched than stranding in the less advanced varieties also receives support from the fact that Phase III overuses pied-piped prepositions in environments that are very strongly associated with stranding in the most advanced varieties. For example, the analysis found several pied-piped *be*-tokens involving adjunct PP chunks in Stage III, whereas *be*-constructions are an almost categorical stranding environment in Stage V (cf. 5.2.2.2). In addition, the data of the less advanced varieties contained more pied-piped free RC tokens with lexicalized adjunct PP

chunks such as *from wherever*. Moreover, the chunk *from where*, which typically appears in constructions with movement PPs such as *come from a place*, turned out to be significantly associated with Phase III in both interrogatives (cf. 5.3.2) and finite RCs (cf. 5.3.3), whereas only the stranded interrogative string *where...from* emerged as being significantly associated with Phase V (cf. 5.3.2). This indicates that, even in the case of high-frequency verb-preposition strings such as *come from*, earlier Dynamic Model stages may prefer pied-piping due to the entrenchment of partly schematic pied-piping constructions. Since pied-piping is more deeply entrenched than stranding in these varieties, pied-piped constructions are much more likely to be activated than structures with stranded prepositions. A similar suggestion was made by Jach (2021: 361), who concluded, based on his corpus-based study of preposition placement in learner Englishes, that, “when competing for activation with stranding constructions, lower-level fronting prototypes outpace stranding competitors at higher levels”. Also note that the fact that *from where* turned out to be significantly associated with Phase III not only in interrogatives but also in finite RCs (and that the free RC string *from wherever* surfaced 6 times in the Phase-III data) suggests that speakers establish links between e.g. the pied-piped WH-interrogative and the pied-piped finite WH-relative clause construction. Consequently, these HCFA results provide support for the idea that constructions at the same level of abstraction are horizontally related in speakers’ mental grammars (cf. 3.1.1.1).

Now, the fact that processing constraints lead to a stronger pied-piping preference in lower stages of evolution might be taken as an indication that varieties become more native-like as they evolve along Schneider’s cycle. For instance, the effects plot which visualizes the interaction of PP_TYPE and VARIETY_TYPE (figure 5.12) clearly suggests that the impact of processing factors should decrease in higher developmental stages. In other words, according to the logistic regression results, varieties should develop a stronger stranding preference with adjunct PPs as they proceed along the Dynamic Model. Importantly, however, the statistical analysis showed that the strong pied-piping preference in varieties at earlier stages leads to innovative meso-constructions that can be said to be entrenched in the construction networks of these varieties. For instance, the use of the chunk *from where*, which turned out to be significantly associated with Phase III in interrogatives (cf. 5.3.2) and finite RCs (cf. 5.3.3), is a characteristic feature of varieties at lower stages of evolution which is conventionalized in the respective speech communities. Consequently, partly schematic *from-where* constructions can be expected to remain part of the construction networks of

these varieties even if these PCEs proceed into higher evolutionary stages. Moreover, in line with hypothesis VI, the analysis showed that PCEs exhibit qualitative innovations that can be attributed to a preference for more transparent and explicit variants. Although the majority of these innovations are low-frequency phenomena produced by individual speakers, at least some innovations, such as the emergence of a *how...look like* meso-construction, can be said to be conventionalized in the Phase-III varieties (cf. 5.3.2). Moreover, the analysis of the missing tokens in section 5.5 showed that a partly schematic *what...talk* construction in which the preposition *about* is omitted is used quite frequently in several New Englishes. The construction of the less advanced varieties thus contains innovative meso-constructions that can be attributed to second-language learning principles. Due to these innovations, the construction networks of PCEs are very unlikely to become identical to those of native varieties even if the former Englishes proceed into higher model phases. Also note that, even though figures 5.34 and 5.35, which visualize the preposition placement families of Phases III and Phase V, imply that all varieties of the same phase have entrenched the same kind of macro-constructions, this does not necessarily mean that these varieties also store the same micro- and meso-constructions. Although processing factors and stylistic variables seem to affect all varieties of the same evolutionary stage in very similar ways, the HCFAs showed that the individual varieties differ from each other with respect to the entrenchment of more substantive constructions, such as V + P or ANTECEDENT + P + WH patterns. However, as the HCFAs conducted in this study were based on small sample sizes, most likely, they did not identify all entrenched meso-constructions that are associated with individual varieties. Further studies with larger sample sizes are thus needed to shed more light on cross-varietal differences at a lower level of abstraction.

Finally, in parts, the results also confirmed hypothesis VII, which assumed that cross-linguistic influence should lead to quantitative innovations in PCEs. First of all, considering that HKE, IndE, PhilE, KenE and TanE only have local languages with pied-piped or in-situ WH-interrogative constructions, cross-linguistic influence may favor pied-piping in WH-interrogatives in some New Englishes. Importantly, though, stranded interrogative constructions exist in major L1s of NigE, and speakers of JamE only receive input for stranded interrogatives in Jamaican Creole. Nevertheless, NigE and JamE also exhibit a weaker preference for stranded interrogative constructions than Phase V. Moreover, the almost categorical pied-piping preference with adjunct PPs in the less advanced varieties cannot be reduced to cross-linguistic influence. Instead, as pointed out above, the fact that the strength of the stranding preference with both WH-interrogatives

and adjunct PPs turned out to be correlated with Dynamic Model stages (cf. 5.2.3.2) can be explained more convincingly by the assumption that processing factors have a stronger effect on less advanced varieties. However, cross-linguistic influence may account for two other effects, namely the stronger preference of SgE for stranded finite WH-relative clause constructions, which may be due to the absence of pied-piped relative clause constructions in the local languages of Singapore (cf. 5.2.3.2), and the strong pied-piping preference of IndE, which stood out from all other varieties. Not only did the logistic regression analysis reveal that IndE favors pied-piping across all PP and clause types (the only exception being free RCs and interrogatives, in which speakers of IndE are equally likely to use stranded and pied-piped prepositions), but also IndE was the only variety (apart from TanE and KenE, which, however, have fewer informal tokens, in which stranded prepositions are most likely to surface) for which the frequencies of stranding tokens were so low that the HCFA did not identify a single entrenched V + P combination (cf. 5.3.1). While this strong pied-piping preference of IndE may, at least in parts, be due to the formal style of IndE (cf. 5.2.3.2), it could also result from cross-linguistic influence. Although there are a few free postpositions in the Dravidian languages, the local languages spoken in India typically have bound postpositions, which do not license stranding as they can normally not be separated from their complements (cf. 3.4). Consequently, speakers of IndE may strongly favor pied-piping when speaking English.

6. General conclusion

The present analysis explored the constructional competition between preposition stranding and pied-piping in 12 varieties of English at different stages of Schneider's Dynamic Model. Adopting a combination of Usage-based Construction Grammar and New Englishes approaches, the study analyzed 10,079 tokens from the ICE corpora to shed more light on the cognitive principles underlying the evolution of PCEs. Three statistical methods were used to gain a full picture of the networks of preposition placement constructions of varieties of English worldwide. In the light of Hoffmann's (2011) findings on preposition placement in BrE and KenE, generalized linear mixed-effects models were fitted to identify differences between varieties and variety phases with respect to processing- and formality-related variables. In addition, two statistical methods were applied to test Hoffmann's (2021) Dynamic Model Productivity hypothesis. The claim that varieties at lower Dynamic Model stages should rely more "on prototypical and frequent fillers" (Hoffmann 2021: 21) was assessed with the help of HCFAs which revealed significant associations between fillers of constructional slots, and individual varieties and variety types. Moreover, LNRE models were used to test if there is a correlation between the productivity of constructional slots and the evolutionary state of a variety. Section 6.1 of this general conclusion will summarize the main findings with respect to the research questions presented in 1.3. Section 6.2 will outline the methodological implications of the present analysis for the study of World Englishes, while 6.3 will show promising avenues for future research.

6.1 Main findings

In the light of previous studies (Hoffmann 2011, 2019, 2021; Brunner & Hoffmann 2020), this project set out to answer the following research questions (cf. 1.3):

- R1. Which factors influence the choice between preposition stranding and pied-piping in Englishes around the world?
- R2. Do processing factors have a stronger effect in L2 Englishes?
- R3. Is there a correlation between the strength of stylistic effects and Dynamic Model stages?
- R4. Does the evolutionary state of the New English varieties correlate with the productivity of preposition stranding and pied-piping constructions?

- R5. Can the use of non-standard preposition placement constructions be linked to processing factors and general principles of L2 learning?
- R6. Does cross-linguistic influence lead to quantitative differences between varieties with respect to the use of preposition placement constructions?

Regarding R1, the results of the corpus study suggest that the constructional competition between preposition stranding and pied-piping is mainly driven by processing constraints and stylistic variables. While there are some idiosyncratic effects of individual prepositions and filler types, speakers generally prefer pied-piping in more formal contexts, while stranding is associated with informal registers. Moreover, all speakers tend to prefer constructions that are easy to process (cf. 5.2). At the same time, these two factors also lead to differences between the construction networks of higher and lower Dynamic Model stages. With respect to R2, it can be said that processing constraints indeed play a greater role in L2 Englishes, which have a stronger preference for pied-piping with adjunct WH-interrogatives, i.e. a construction type that involves a reduction in processing cost. Since speakers of varieties at earlier stages receive less input, their construction networks are not as deeply entrenched as those of higher Dynamic Model stages. As a result, speakers of less advanced varieties are more strongly affected by processing constraints during online production (see Hoffmann 2011: 275). Importantly, however, this effect only applies to the use of WH-interrogative constructions but not to finite WH-relative clauses, which involve more processing cost (cf. 5.2, 5.3.2). The results thus indicate that processing factors only have a stronger effect in L2 Englishes in contexts in which the processing load is not maximal. In addition, turning to R3, the study found a correlation between the strength of stylistic variables and Dynamic Model stages, with more advanced varieties exhibiting stronger formality effects (cf. 5.2.2.2, 5.2.4.2). While the construction networks of the more advanced varieties contain stranded and pied-piped WH-relative clause constructions that are associated with less formal and more formal contexts respectively, the less advanced varieties always prefer the pied-piping construction, i.e. the variant that is not only easier to process but also more frequent in the formal input that is typical of lower Dynamic Model stages.

Coming to R4, the productivity of the slots of stranding constructions, which tend to be very idiomatic, indeed turned out to be correlated with Dynamic Model stages, with Phase V displaying the highest and Phase III displaying the lowest productivity (cf. 5.4.1, 5.4.2). Particularly when using non-prototypical stranding constructions such as stranded finite WH-relative clauses, speakers of less advanced varieties rely more on entrenched

meso-constructions than speakers of more advanced varieties (cf. 5.3.1). In contrast, this effect was not confirmed for the slots of pied-piped finite RCs. Instead, P + WH combinations in pied-piped finite WH-relative clause constructions turned out to be most productive in the least advanced varieties (cf. 5.4.3), which can be explained by the fact that, due to the weaker entrenchment of constructions and a preference for explicit and transparent structures, varieties at lower stages of evolution tend to use prepositions in pied-piped finite WH-relatives in innovative ways (R5). Other innovations that surfaced in the PCE data, such as doubling, missing or extra tokens, can also be said to result from general second-language learning principles or processing factors (cf. 5.5). Finally, some of the quantitative effects found in the present analysis, such as the strong preference for pied-piped interrogative constructions in some Phase-III varieties and the extremely strong pied-piping probability of IndE, may be due to cross-linguistic influence (R6; cf. 5.2.3.2).

In sum, the findings of the corpus study suggest that differences between the construction networks of more and less advanced varieties result from a complex interaction of processing factors, input effects, second-language learning principles and cross-linguistic influence.

6.2 Methodological implications

The present analysis has several methodological implications for the study of World Englishes. First of all, the results are strikingly similar to those of Hoffmann (2011), who investigated only two varieties, viz. KenE and BrE. Not only does the finding that less advanced varieties have a stronger preference for pied-piping with adjunct PPs confirm Hoffmann's (2011: 270) claim that L2 speakers have a stronger preference for prototypical constructions that are easy to process, but also the results indicate that the lack of register distinctions which Hoffmann (2011: 155, 167) observed in KenE is a general characteristic of the Phase-III varieties. Even at a lower level of abstraction, the results are largely in line with Hoffmann's (2011) findings. In accordance with Hoffmann's (2011: 137–138) observation that the idiomatic string *what...be like* is more deeply entrenched in BrE than in KenE, the present study found that a partly schematic *what...like* construction is deeply entrenched in the construction networks of Phase V, while it occurs with significantly lower frequencies in the less advanced varieties. Moreover, Hoffmann's (2011: 145) finding that BrE has entrenched the stranded frequency PP string *how long...for*, while KenE prefers the pied-piped variant *for how*

long turned out to apply to Phases V and III in general. Likewise, the weaker entrenchment of stranded *be*-constructions which Hoffmann (2011: 240) observed in KenE was confirmed for all Stage-III varieties in this analysis. Hoffmann's (2011) results for BrE and KenE thus largely correspond to those for Phases V and III in the present study. Furthermore, in the logistic regression analysis in chapter 5, the variable VARIETY_TYPE turned out to be a reliable proxy for the factor VARIETY as the variety-specific effects identified in 5.2.3 largely confirmed the effects of VARIETY_TYPE found in 5.2.2 (the only exception being the extremely strong pied-piping preference of IndE). This shows that varieties subsumed under the same Dynamic Model phase indeed pattern very similarly in terms of their constructional choices. Although Schneider's (2007: 113–250) categorization of varieties according to Dynamic Model stages is not based on empirical studies and may thus fail to take into account input differences that might exist between varieties that are assigned to the same model stage (Mußermann 2024: 23), this suggests that, at least with respect to processing factors and formality effects, it is possible to obtain valid results using a phase-related approach.

Besides this, the study illustrated the importance of combining statistical techniques that focus on different levels of abstraction. For instance, the logistic regression analysis in 5.2.3 suggested that NigE and HKE have a relatively strong preference for stranded finite RCs, which might be taken as an indication that these varieties have more abstract representations of a stranded finite WH-relative clause construction. However, the HCFA in 5.3.1 revealed that this stronger preference for stranded finite RCs in some Stage-III varieties can be explained by the fact that these varieties rely more on lexicalized verb-preposition strings. This underlines that it is useful to complement studies that focus on macro-constructions with statistical methods that analyze cross-varietal differences at a lower level of abstraction. Moreover, as pointed out in 5.6, the results of studies that only focus on abstract constructions may be mistaken as implying that varieties become more native-like as they evolve along Schneider's cycle. Since linguistic innovations in PCEs typically emerge at the meso-constructional level, differences between more and less advanced varieties can frequently only be identified with the help of statistical techniques that focus on lower levels of schematicity and take into account more item-based variation. The present study thus illustrated how a combination of several statistical techniques, namely logistic regression modeling, HCFA, and LNRE modeling, can be applied to identify cross-varietal differences at all levels of abstraction, which makes it possible to give a full account of the construction networks of different variety types.

Finally, the present study showed that topic modeling can be useful to investigate formality effects in corpus-based research. As topic modeling offers an objective technique to identify topic and formality differences between corpora, it can be applied to implement a nuanced formality classification that is not only based on text types but also on topics. Although the ICE corpora follow a common corpus design, topic differences turned out to be of major importance in the present study. For instance, in 5.2.2.1, it was shown that the private correspondence texts in ICE-IND are much more formal than those of the other ICE components as they tend to focus on administrative issues, whereas the private correspondence texts of the other varieties mainly deal with personal matters. Topic modeling thus allows researchers to take into account cross-varietal differences regarding topic and formality that are likely to exist even in cases in which all corpora share the same categories. Moreover, particularly studies that analyze data from several corpora which differ from each other in terms of their design can benefit from the use of topic models. Also note that topic modeling is a promising technique for studies in the field of Usage-based Construction Grammar, which assume that the entrenchment of constructions is directly linked to input. For example, the present analysis demonstrated that the distinction between personal and non-personal topics only affects speakers' constructional choices in Phases IV and V but not in Phase III. This effect can be explained by the fact that speakers of less advanced varieties typically do not encounter input for English in connection with private matters and personal topics, which means that these speakers store fewer pragmatic associations between constructions and informal usage situations (cf. 5.2.4.2). Topic modeling thus offers a useful tool for studies that aim to explore cross-varietal differences which can be attributed to input effects.

6.3 Avenues for future research

A major limitation of the present study was that several interactions of variables were too complex to be analyzed with the help of generalized linear mixed-effects models. For instance, the three-way interaction of clause type, PP type and variety/variety type would have introduced many low frequency cells, which would have led to overfitting. As a result, the logistic regression analysis could not test whether the strong pied-piping preference with adjunct PPs in the earlier Dynamic Model phases indeed mostly applies to WH-interrogative constructions. Although the claim that less advanced varieties are more likely to use pied-piped adjunct interrogative constructions than more advanced

varieties receives strong support from the logistic regression models which indicated that Phase-III varieties have a stronger pied-piping preference with both adjunct PPs and WH-interrogatives (cf. 5.2.2.2 and 5.2.3.2), and from the HCFA which identified several P + FILLER combinations for adjunct interrogatives that are significantly associated with Phase III (cf. 5.3.2), the assumed interaction effect of PP type, clause type and variety type definitely needs to be corroborated by future research using larger corpora. A larger database would also make it possible to take a closer look at preposition placement in *it*-cleft and exclamative constructions, whose token numbers in the ICE data were too low to be included in the statistical analysis. As many of the HCFAs in section 5.3 are likely to suffer from high Type II error rates, larger sample sizes would also be useful to replicate the HCFAs of the present study in order to identify more significant associations between fillers of constructional slots and individual varieties.

Furthermore, the results suggested that cross-varietal differences with respect to processing constraints typically arise in contexts in which the processing load is not maximal, i.e. in WH-interrogatives but not in finite RCs. In order to investigate whether the complexity of a construction plays a role in the emergence of processing-related differences between L1 and L2 varieties, experimental studies could be useful that explore the acceptability of stranded and pied-piped WH-interrogative and finite RC constructions across several L1 and L2 varieties. Besides this, the analysis indicated that processing-related differences between L1 and L2 Englishes are most likely to surface in studies that investigate competing constructions which are semantically interchangeable and which are strongly affected by variables that have an impact on parsing efficiency. In order to assess the validity of this claim, other alternation phenomena should be analyzed across a wide range of PCEs and learner Englishes. Even though the dative alternation, the genitive alternation, particle placement or *that*-complementation have been studied quite extensively in World Englishes (cf. 2.4.1.1.), apart from Bernaisch, Gries & Mukherjee (2014), who studied the dative alternation with the verb *give*, these investigations focused exclusively on abstract constructions. However, the present analysis demonstrated that cross-varietal differences that can be attributed to processing factors frequently surface at the meso-constructional level (recall e.g. that, due to processing reasons, Phase III prefers the *for how long* construction, while Phase V has an entrenched *how long...for* construction; cf. 5.3.2). Alternation phenomena of varying degrees of complexity should thus be investigated with a focus on lower-level constructions in order to shed more light on the role of processing factors and semantic variables in varieties at different evolutionary stages.

Finally, the present study showed that processing factors interact in complex ways with stylistic variables and input effects. At the same time, the size of the ICE corpora turned out to be too low for a thorough analysis of formality effects in the individual varieties. Now, the Corpus of Global Web-based English can be expected to offer much higher token frequencies for all types of stranding and pied-piping constructions. However, this corpus is not an ideal database either as it only includes written language. A more detailed investigation of the interplay of stylistic variables and processing factors in World Englishes thus calls for the compilation of larger corpora that include more data from informal, spoken contexts, and metadata on the mother tongue and additional languages of the corpus speakers. Such a database would open up new opportunities to systematically investigate the various factors that turned out to affect speakers' constructional choices in the present study – processing factors, input effects, second-language learning principles and cross-linguistic influence – across different constructions and varieties in order to further deepen our understanding of how “alternate ways of saying ‘the same’ thing” (Labov 1972: 188) are represented in the mental grammars of speakers of varieties of English worldwide.

Appendix

A.1: The 20 most likely terms for the 20 topics included in the present study

TOPIC	national development	situational description	administration
most likely terms	develop countri area nation plan govern industri programm problem econom project product popul import resourc local environ communiti region polici	back hand room eye turn head door left car face light walk hous open night stop day thought road long	offic letter meet date depart inform dear receiv member applic staff account form copi request servic requir issu sincer book
TOPIC	society	court	technology
most likely terms	women famili children man life live god mother men woman father child young home love parent girl person church work	case court sir polic evid person accus question law fact lord statement act offic charg wit number mister matter told	system comput design oper inform program model process data set problem work requir technolog machin applic servic network power product

A.1 (cont.)

TOPIC	education	economy	news
most likely terms	school	year	news
	student	compani	war
	univers	market	countri
	educ	busi	nation
	year	dollar	report
	teacher	bank	unit
	studi	price	presid
	teach	pay	minist
	work	cost	india
	class	rate	forc
	colleg	employ	state
	learn	money	today
	train	cent	militari
	level	increas	hundr
	research	million	peac
	scienc	govern	yesterday
	graduat	tax	secur
	lectur	hundr	leader
	academ	percent	world
	programm	thousand	twenti
TOPIC	art and literature	disease	language
most likely terms	music	health	languag
	stori	patient	english
	play	doctor	mean
	art	hospit	question
	work	medic	thing
	book	diseas	learn
	write	case	write
	paint	year	speak
	song	drug	understand
	artist	blood	talk
	danc	bodi	kind
	film	care	point
	world	age	read
	show	treatment	communic
	perform	effect	problem
	great	problem	make
	life	test	find
	charact	risk	studi
	record	children	group
	read	control	term

A.1 (cont.)

TOPIC	personal	government	science
most likely terms	thing good talk back day work kind week start bit call put thought nice rememb year friend make suppos guy	govern parti state member elect polit presid minist nation speaker public issu law committe vote hous nigeria council countri constitut	water result form studi effect increas sampl materi high area measur surfac cell concentr method plant air product number chemic
TOPIC	philosophy	Asia	history and culture
most likely terms	social societi cultur human polit view power role world natur individu chang state exist relat term tradi import sens group	hong kong singapor chines year china philippin dollar filipino group nineti citi japanes japan manila open foreign month peso lee	year jamaica island zealand citi ireland area canada road north maori northern john park canadian irish street hous communiti jamaican

A.1 (cont.)

TOPIC	household and gardening	sports
most likely terms	food	back
	water	play
	plant	good
	tree	game
	make	ball
	fish	run
	cook	point
	colour	team
	anim	side
	eat	win
	cut	number
	oil	line
	bird	half
	leav	minut
	put	player
	small	twenti
	flower	final
	add	race
	forest	goal
	green	forward

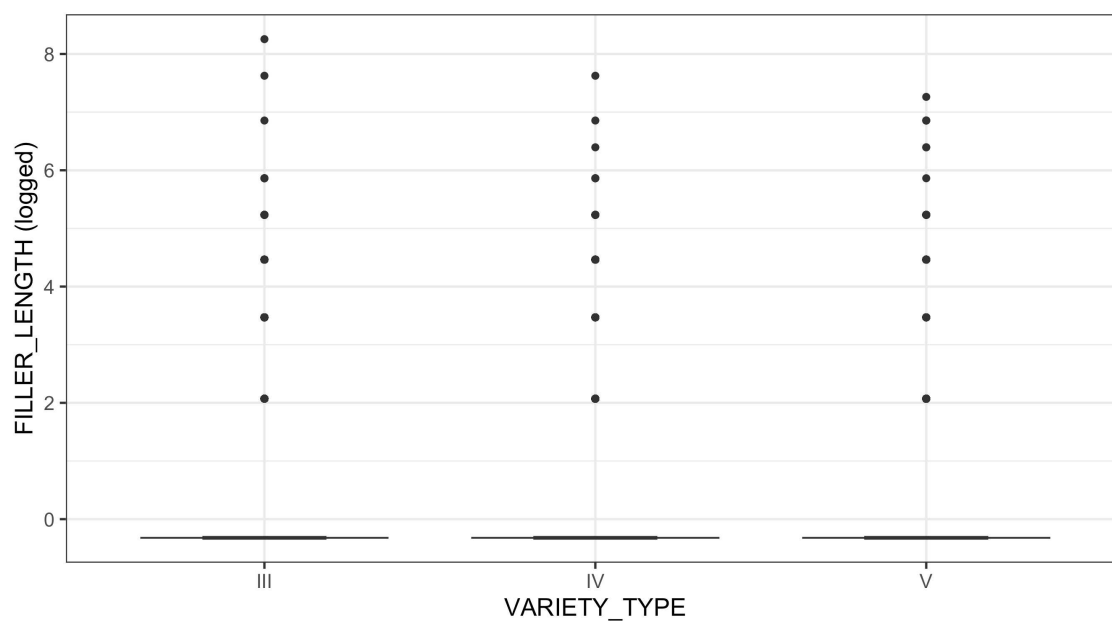
A.2: *CLAUSE_TYPE * VOICE * VARIETY_TYPE* (absolute frequencies)

CLAUSE_TYPE	VOICE	VARIETY_TYPE	PREPOSITION_ PLACEMENT	Abs. freq.
embedded interrogative	active	III	stranding	260
			pied-piping	81
		IV	stranding	176
			pied-piping	13
		V	stranding	313
			pied-piping	28
	passive	III	stranding	11
			pied-piping	22
		IV	stranding	3
			pied-piping	4
		V	stranding	14
			pied-piping	12
finite RC	active	III	stranding	251
			pied-piping	1,596
		IV	stranding	144
			pied-piping	565
		V	stranding	234
			pied-piping	1,262
	passive	III	stranding	12
			pied-piping	615
		IV	stranding	10
			pied-piping	241
		V	stranding	9
			pied-piping	510
free RC	active	III	stranding	441
			pied-piping	34
		IV	stranding	247
			pied-piping	8
		V	stranding	543
			pied-piping	8
	passive	III	stranding	19
			pied-piping	3
		IV	stranding	5
			pied-piping	0
		V	stranding	19
			pied-piping	0

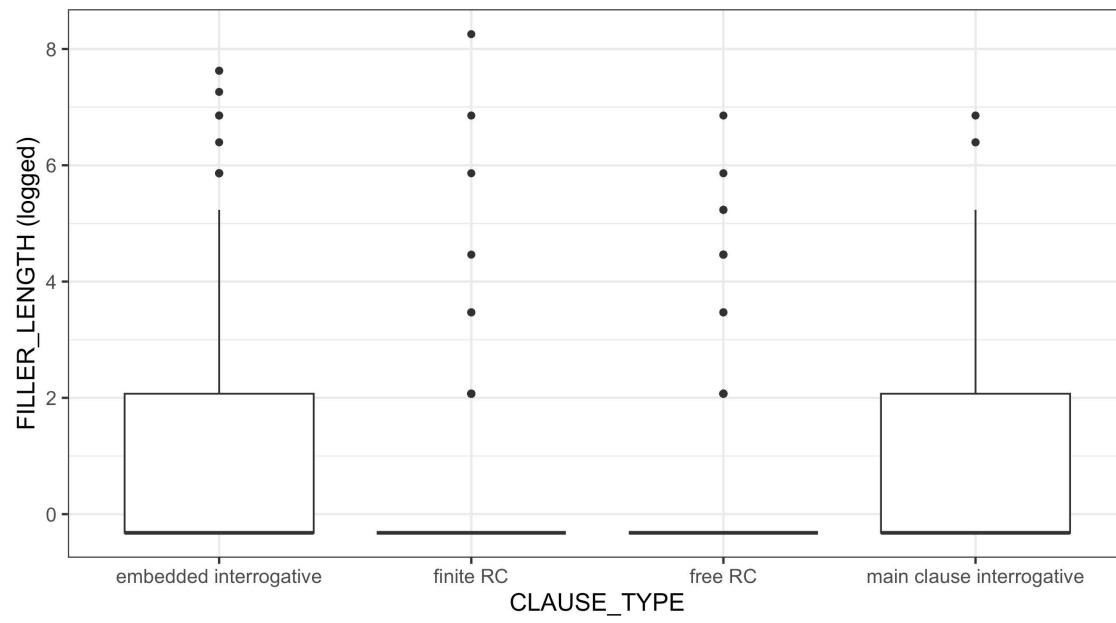
A.2 (cont.)

CLAUSE_TYPE	VOICE	VARIETY_TYPE	PREPOSITION_ PLACEMENT	Abs. freq.
main clause interrogative	active	III	stranding	332
			pied-piping	127
		IV	stranding	214
			pied-piping	33
		V	stranding	462
			pied-piping	32
	passive	III	stranding	13
			pied-piping	14
		IV	stranding	6
			pied-piping	7
		V	stranding	10
			pied-piping	11

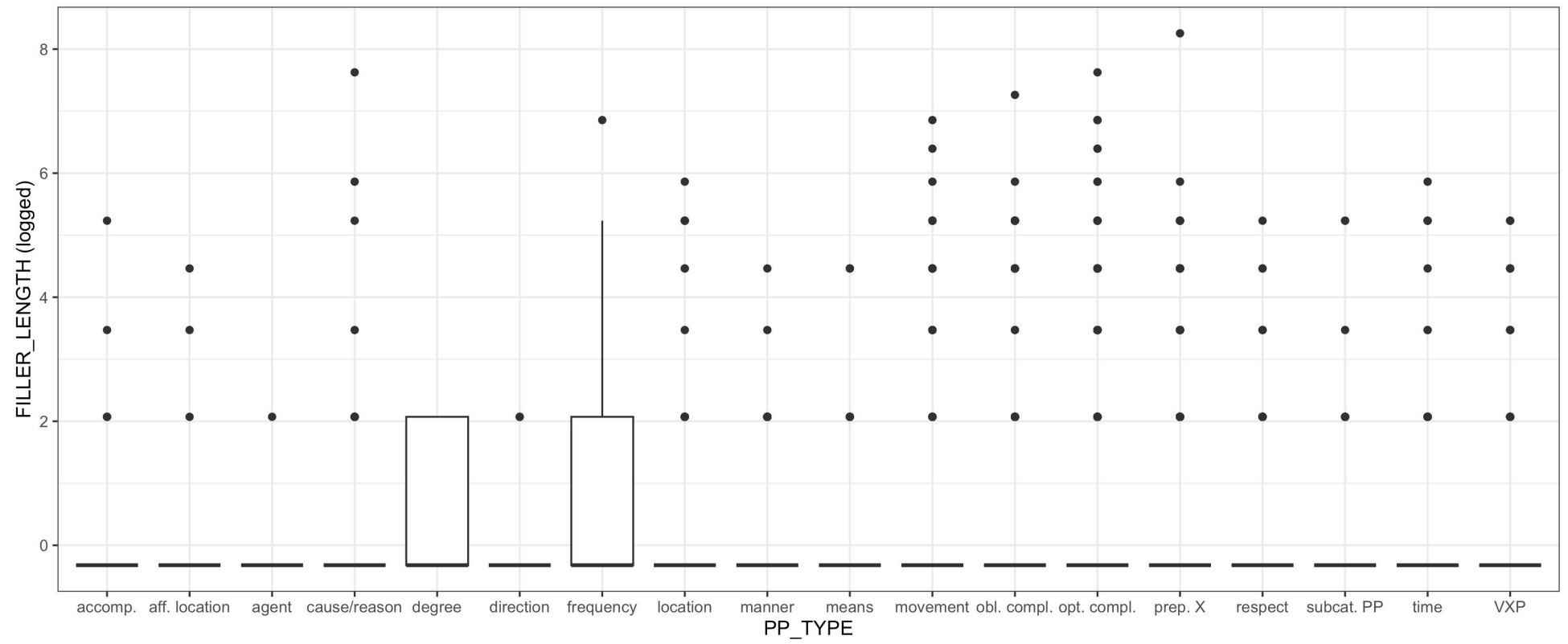
A.3: Boxplot: *FILLER_LENGTH* (logged) * *VARIETY_TYPE*



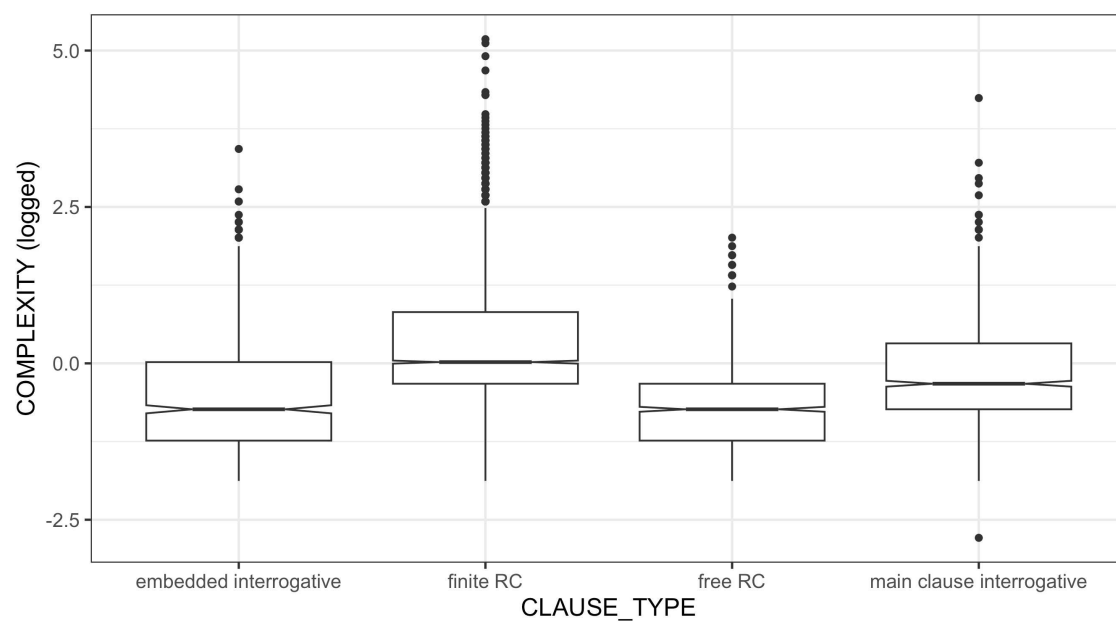
A.4: Boxplot: *FILLER_LENGTH (logged) * CLAUSE_TYPE*



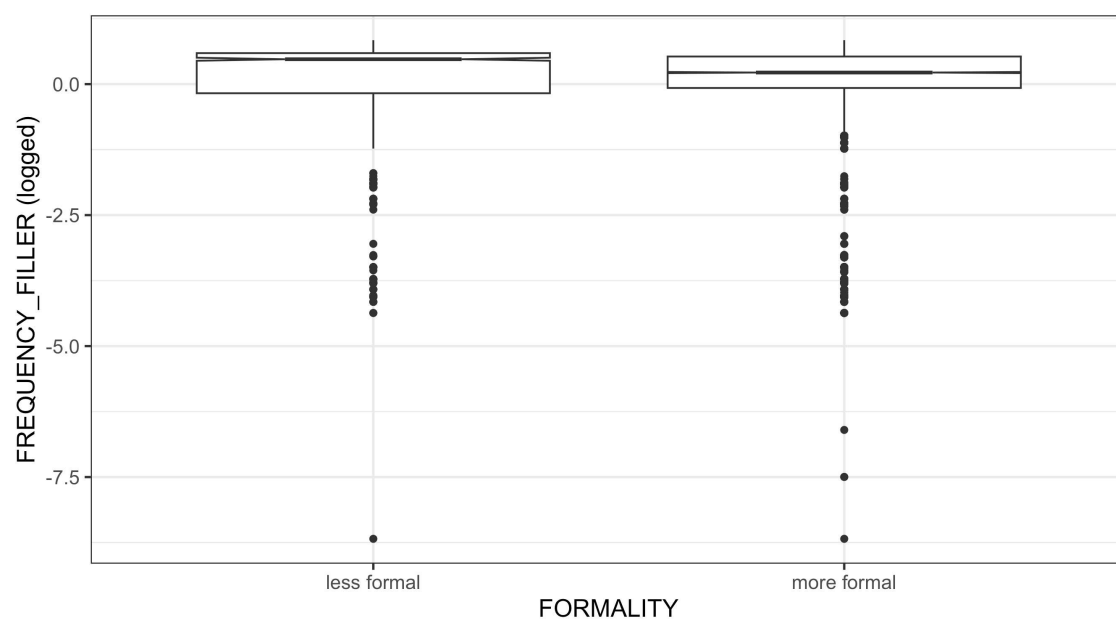
A.5: Boxplot: *FILLER_LENGTH (logged) * PP_TYPE*



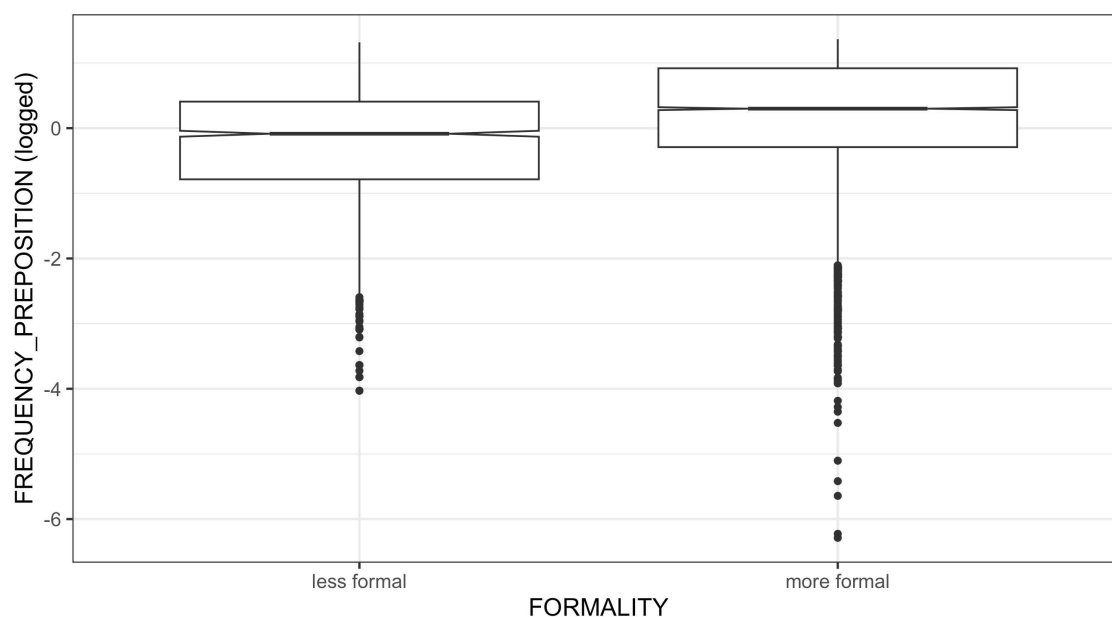
A.6: Boxplot: COMPLEXITY (logged) * CLAUSE_TYPE



A.7: Boxplot: FREQUENCY_FILLER (logged) * FORMALITY



A.8: Boxplot: *FREQUENCY_PREPOSITION (logged) * FORMALITY*



A.9: Top 3 fillers per ICE variety (absolute and relative frequencies)

Variety	Filler	Abs. freq.	Rel. freq.
BrE	which	4,231	28.50%
	what	4,041	27.22%
	who	2,246	15.13%
CanE	what	4,204	33.05%
	who	2,104	16.54%
	which	2,094	16.46%
IrE	what	4,117	30.99%
	which	2,884	21.71%
	who	2,169	16.33%
NZE	what	4,209	30.19%
	which	2,971	21.31%
	who	2,261	16.22%
JamE	what	4,570	29.66%
	which	3,067	19.91%
	who	2,596	16.85%
SgE	what	4,491	31.05%
	which	2,717	18.78%
	how	2,058	14.23%
KenE	what	2,874	23.15%
	which	2,771	22.32%
	who	2,751	22.16%

A.9 (cont.)

Variety	Filler	Abs. freq.	Rel. freq.
TanE	which	3,237	37.63%
	who	1,609	18.70%
	what	1,455	16.91%
HKE	what	3,787	29.35%
	which	2,598	20.14%
	how	2,034	15.77%
IndE	what	3,970	27.96%
	which	3,868	27.24%
	who	1,867	13.15%
NigE	what	4,006	29.88%
	who	2,328	17.36%
	which	2,240	16.71%
PhilE	what	5,025	32.78%
	which	2,501	16.31%
	who	2,364	15.42%

A.10: HCFA results for VARIETY_TYPE/VARIETY * PREPOSITION_PLACEMENT * CLAUSE_TYPE of BE-tokens (only significant results are shown)

VARIETY_ TYPE	VARIETY	PREPOSITION_ PLACEMENT	CLAUSE_TYPE	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
.	.	pied-piping	.	36	534	464.427	<	5.17978743914047E-255	***	0.933
.	.	stranding	.	1032	534	464.427	>	1.03595748782809E-254	***	0.933
III	.	pied-piping	.	28	10.1124	31.641	>	1.43361637622153E-05	***	0.017
V	.	pied-piping	.	5	18.2022	9.5757	<	0.00128118423465	**	0.013
.	IndE	pied-piping	.	8	1.382	31.6917	>	0.00230086651926	**	0.006
.	TanE	pied-piping	.	5	0.809	21.7113	>	0.033853643387126	*	0.004
.	.	pied-piping	finite RC	16	0.9438	240.1877	>	7.063269E-14	***	0.014
.	.	stranding	finite RC	12	27.0562	8.3785	<	0.007898151960227	**	0.014
III	.	pied-piping	finite RC	12	0.2651	519.4563	>	5.57309567454983E-15	***	0.011
IV	.	pied-piping	finite RC	3	0.2015	38.8665	>	0.03393649541575	*	0.003
.	IndE	pied-piping	embedded interrogative	5	0.4451	46.6123	>	0.011893082656441	*	0.004
.	Phile	pied-piping	finite RC	4	0.0795	193.3374	>	0.000186756341485	***	0.004
.	JamE	pied-piping	finite RC	3	0.1228	67.4127	>	0.032887805175216	*	0.003
.	TanE	pied-piping	finite RC	2	0.0212	184.7004	>	0.026144468542436	*	0.002

*A.11: VARIETY * FORMALITY * CLAUSE_TYPE (absolute and relative frequencies)*

VARIETY	FORMALITY	CLAUSE_TYPE	PREPOSITION_ PLACEMENT	Abs. freq.	Rel. freq.
CanE	less formal	freeRC_interrogative	stranding	158	98.14%
			pied-piping	3	1.86%
	more formal	finiteRC	stranding	15	46.88%
			pied-piping	17	53.12%
		freeRC_interrogative	stranding	174	98.86%
			pied-piping	2	1.14%
BrE	less formal	finiteRC	stranding	22	11.11%
			pied-piping	176	88.89%
	more formal	freeRC_interrogative	stranding	143	97.28%
			pied-piping	4	2.72%
		finiteRC	stranding	32	69.57%
			pied-piping	14	30.43%
IrE	less formal	freeRC_interrogative	stranding	165	92.18%
			pied-piping	14	7.82%
	more formal	finiteRC	stranding	37	7.91%
			pied-piping	431	92.09%
	less formal	freeRC_interrogative	stranding	178	98.89%
			pied-piping	2	1.11%
NZE	less formal	finiteRC	stranding	11	64.71%
			pied-piping	6	35.29%
	more formal	freeRC_interrogative	stranding	149	93.12%
			pied-piping	11	6.88%
		finiteRC	stranding	45	15.20%
			pied-piping	251	84.80%
NZE	less formal	freeRC_interrogative	stranding	167	98.24%
			pied-piping	3	1.76%
	more formal	finiteRC	stranding	16	80.00%
			pied-piping	4	20.00%
		freeRC_interrogative	stranding	224	96.55%
			pied-piping	8	3.45%
	more formal	finiteRC	stranding	63	25.40%
			pied-piping	185	74.60%

VARIETY	FORMALITY	CLAUSE_TYPE	PREPOSITION_ PLACEMENT	Abs. freq.	Rel. freq.
JamE	less formal	freeRC_interrogative	stranding	160	96.39%
			pied-piping	6	3.61%
	more formal	finiteRC	stranding	28	65.12%
			pied-piping	15	34.88%
		freeRC_interrogative	stranding	201	89.73%
			pied-piping	23	10.27%
SgE	less formal	finiteRC	stranding	57	14.07%
			pied-piping	348	85.93%
	more formal	freeRC_interrogative	stranding	115	99.14%
			pied-piping	1	0.86%
		finiteRC	stranding	12	66.67%
			pied-piping	6	33.33%
KenE	less formal	freeRC_interrogative	stranding	175	92.11%
			pied-piping	15	7.89%
	more formal	finiteRC	stranding	55	25.94%
			pied-piping	157	74.06%
		freeRC_interrogative	stranding	34	89.47%
			pied-piping	4	10.53%
TanE	less formal	finiteRC	stranding	9	47.37%
			pied-piping	10	52.63%
	more formal	freeRC_interrogative	stranding	173	85.64%
			pied-piping	29	14.36%
		finiteRC	stranding	40	15.75%
			pied-piping	214	84.25%
TanE	less formal	freeRC_interrogative	stranding	0	0.00%
			pied-piping	1	100.00%
	more formal	finiteRC	stranding	1	50.00%
			pied-piping	1	50.00%
		freeRC_interrogative	stranding	57	85.07%
			pied-piping	10	14.93%
	more formal	finiteRC	stranding	30	12.93%
			pied-piping	202	87.07%

VARIETY	FORMALITY	CLAUSE_TYPE	PREPOSITION_ PLACEMENT	Abs. freq.	Rel. freq.
HKE	less formal	freeRC_interrogative	stranding	69	98.57%
			pied-piping	1	1.43%
		finiteRC	stranding	2	66.67%
			pied-piping	1	33.33%
	more formal	freeRC_interrogative	stranding	108	80.60%
			pied-piping	26	19.40%
		finiteRC	stranding	33	19.88%
			pied-piping	133	80.12%
IndE	less formal	freeRC_interrogative	stranding	32	65.31%
			pied-piping	17	34.69%
		finiteRC	stranding	2	7.69%
			pied-piping	24	92.31%
	more formal	freeRC_interrogative	stranding	73	60.83%
			pied-piping	47	39.17%
		finiteRC	stranding	29	6.94%
			pied-piping	389	93.06%
NigE	less formal	freeRC_interrogative	stranding	93	93.00%
			pied-piping	7	7.00%
		finiteRC	stranding	8	38.10%
			pied-piping	13	61.90%
	more formal	freeRC_interrogative	stranding	161	93.60%
			pied-piping	11	6.40%
		finiteRC	stranding	55	21.48%
			pied-piping	201	78.52%
Phile	less formal	freeRC_interrogative	stranding	137	95.80%
			pied-piping	6	4.20%
		finiteRC	stranding	14	66.67%
			pied-piping	7	33.33%
	more formal	freeRC_interrogative	stranding	138	88.46%
			pied-piping	18	11.54%
		finiteRC	stranding	40	16.19%
			pied-piping	207	83.81%

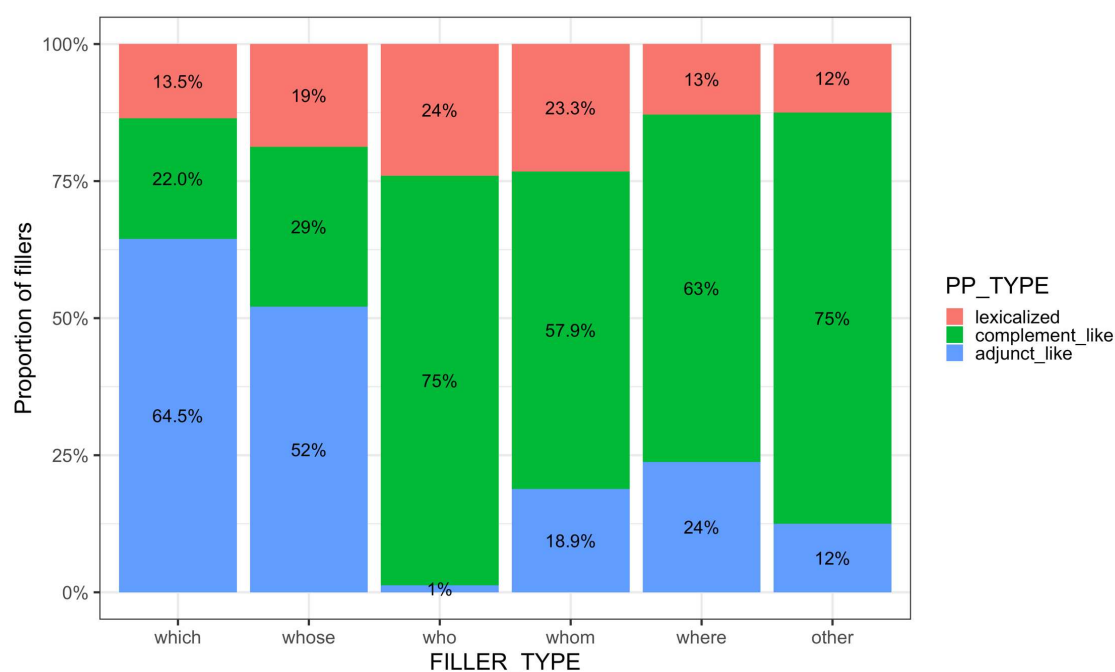
*A.12: VARIETY_TYPE * FILLER_TYPE in finite RCs (absolute and relative frequencies)*

VARIETY_ TYPE	FILLER_TYPE	PREPOSITION_ PLACEMENT	Abs. freq.	Rel. freq.
III	what	stranding	2	66.67%
		pied-piping	1	33.33%
	where	stranding	37	52.86%
		pied-piping	33	47.14%
	which	stranding	184	8.28%
		pied-piping	2,037	91.72%
	who	stranding	20	90.91%
		pied-piping	2	9.09%
	whom	stranding	17	12.78%
		pied-piping	116	87.22%
	whose	stranding	3	12.50%
		pied-piping	21	87.50%
IV	what	stranding	1	100.00%
		pied-piping	0	0.00%
	where	stranding	16	100.00%
		pied-piping	0	0.00%
	which	stranding	99	11.88%
		pied-piping	734	88.12%
	who	stranding	27	96.43%
		pied-piping	1	3.57%
	whom	stranding	11	14.67%
		pied-piping	64	85.33%
	whose	stranding	0	0.00%
		pied-piping	7	100.00%

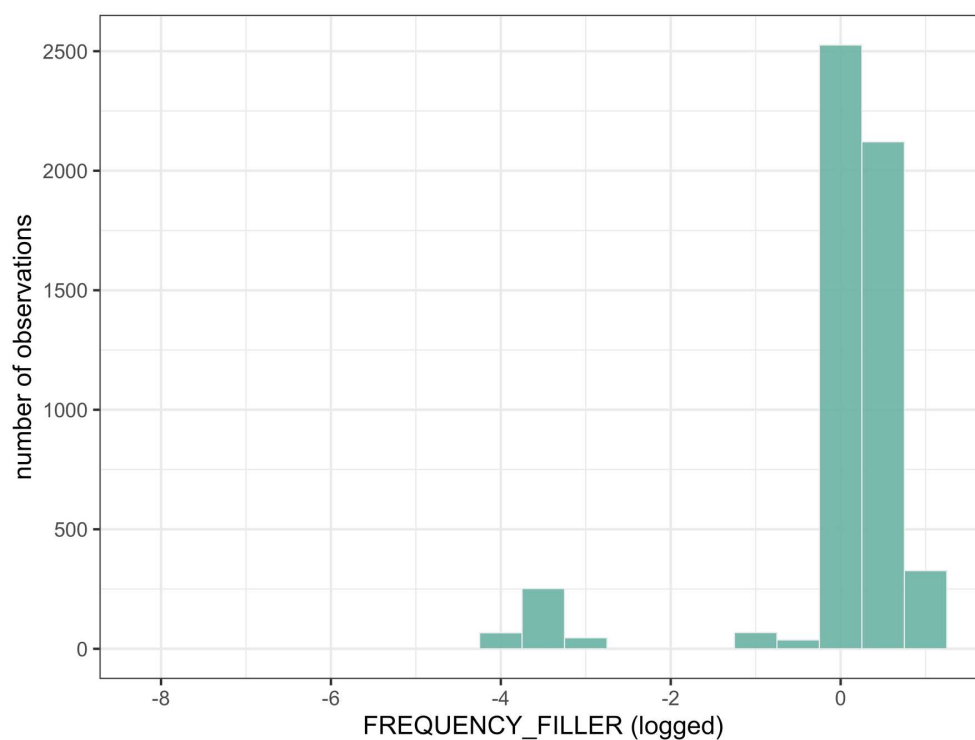
A.12 (cont.)

VARIETY_ TYPE	FILLER_TYPE	PREPOSITION_ PLACEMENT	Abs. freq.	Rel. freq.
V	what	stranding	1	100.00%
		pied-piping	0	0.00%
	whence	stranding	0	0.00%
		pied-piping	3	100.00%
	where	stranding	12	80.00%
		pied-piping	3	20.00%
	which	stranding	190	10.33%
		pied-piping	1,650	89.67%
	who	stranding	28	96.55%
		pied-piping	1	3.45%
	whom	stranding	6	5.45%
		pied-piping	104	94.55%
	whose	stranding	6	35.29%
		pied-piping	11	64.71%

A.13: Distribution of PP_TYPE across FILLER_TYPE



A.14: Frequency histogram for *FREQUENCY_FILLER* (logged) in finite RCs



A.15: *VARIETY_TYPE* * *VOICE* * *TOPIC* (absolute and relative frequencies)

VARIETY_TYPE	VOICE	TOPIC	Abs. freq.	Rel. freq.
III	active	personal_less_formal	45	2.44%
		other_less_formal	71	3.85%
		more_formal	1,730	93.72%
	passive	personal_less_formal	4	0.64%
		other_less_formal	7	1.12%
		more_formal	616	98.25%
IV	active	personal_less_formal	34	4.80%
		other_less_formal	35	4.94%
		more_formal	640	90.27%
	passive	personal_less_formal	1	0.40%
		other_less_formal	6	2.39%
		more_formal	244	97.21%
V	active	personal_less_formal	110	7.35%
		other_less_formal	18	1.20%
		more_formal	1,368	91.44%
	passive	personal_less_formal	8	1.54%
		other_less_formal	7	1.35%
		more_formal	504	97.11%

*A.16: HCFA results for VARIETY_TYPE * V * P in free RCs/interrogatives (only significant results are shown)*

VARIETY_TYPE	V	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
V	be	like	184	68.7411	193.2558	>	5.60834256501896E-27	***	0.032
V	talk	about	132	37.0024	243.8908	>	9.93432229684707E-30	***	0.026
V	come	from	101	13.2835	579.2287	>	6.6779643097573E-49	***	0.024
V	look	like	72	19.9336	135.997	>	4.09867891727612E-15	***	0.014
V	be	to	2	45.8274	41.9147	<	3.4018339E-13	***	0.012
V	look	about	0	29.3098	29.3098	<	5.35322147663E-09	***	0.008
V	come	about	0	24.2463	24.2463	<	8.7856800960752E-07	***	0.007
V	talk	like	0	25.1653	25.1653	<	3.4830321277108E-07	***	0.007
V	look	at	27	6.6076	62.9351	>	6.83679795675268E-05	***	0.006
V	talk	for	1	23.9071	21.9489	<	3.09729772343064E-05	***	0.006
V	get	from	28	5.1214	102.2045	>	5.212105470949E-08	***	0.006
V	talk	from	0	20.2721	20.2721	<	4.78787255954552E-05	***	0.006
V	think	of	21	0.7898	517.159	>	1.98686891794585E-18	***	0.006
V	go	to	27	6.5783	63.3972	>	6.23473640501451E-05	***	0.006
V	be	up to	30	6.7468	80.1434	>	1.26588100510945E-06	***	0.006
V	be	with	4	24.1867	16.8482	<	0.016160889157444	*	0.006
V	come	like	0	16.4899	16.4899	<	0.002141557907369	**	0.005
V	go	about	0	14.5088	14.5088	<	0.01565641232018	*	0.004
V	interested	in	15	0.6279	328.9652	>	1.241741417E-11	***	0.004
V	be	worth	16	2.8006	62.2096	>	0.001566731248436	**	0.004
V	use	for	11	1.5099	59.6477	>	0.01873874367164	*	0.003
V	sound	like	14	1.2583	129.024	>	2.8026129338164E-06	***	0.003
V	work	on	10	0.7383	116.1846	>	0.000216623710936	***	0.003
V	believe	in	9	0.5102	141.2715	>	0.000130468846031	***	0.002
V	focus	on	7	0.2232	205.7573	>	0.000144522324714	***	0.002
V	deal	with	9	0.6291	111.3845	>	0.000773478412298	***	0.002
IV	be	like	94	32.3708	117.3329	>	2.099187E-14	***	0.017
IV	talk	about	73	17.4247	177.2549	>	6.76280702253076E-19	***	0.015
IV	come	from	46	6.2553	252.5284	>	4.45823922064727E-20	***	0.011
IV	be	to	0	21.5805	21.5805	<	1.28429250674123E-05	***	0.006
IV	look	at	21	3.1116	102.8393	>	7.068738949229E-07	***	0.005
IV	go	to	20	3.0978	92.2217	>	4.47564934061265E-06	***	0.005
IV	look	about	0	13.8022	13.8022	<	0.031816545843533	*	0.004
IV	get	from	15	2.4117	65.7069	>	0.00138277207917	**	0.003
IV	refer	to	12	0.9772	124.337	>	2.04465642560985E-05	***	0.003

VARIETY_TYPE	V	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
IV	call	from	8	0.3517	166.325	>	0.000136113750222	***	0.002
IV	go	through	7	0.3442	128.7033	>	0.002696784349743	**	0.002
IV	lead	to	6	0.3119	103.7335	>	0.031423695118816	*	0.002
IV	deal	with	9	0.2963	255.6679	>	1.18705713999272E-06	***	0.002
IV	focus	on	5	0.1051	227.9738	>	0.003149029177808	**	0.001
IV	fall	under	2	0.0012	3329.3345	>	0.021475778811615	*	0.001
III	talk	about	156	27.8303	590.2729	>	9.83981820316621E-60	***	0.035
III	come	from	82	9.9908	519.01	>	1.62149462879131E-41	***	0.02
III	look	like	57	14.9925	117.7009	>	2.7554034E-12	***	0.011
III	be	for	17	49.1165	21.0005	<	0.003107600905036	**	0.009
III	be	to	1	34.4678	32.4968	<	1.05369308237E-09	***	0.009
III	look	about	0	22.0445	22.0445	<	8.05336983449709E-06	***	0.006
III	refer	to	25	1.5607	352.022	>	2.94376980964827E-17	***	0.006
III	come	about	0	18.2361	18.2361	<	0.000370521064472	***	0.005
III	look	at	25	4.9697	80.7318	>	4.32913648781489E-06	***	0.005
III	ask	for	20	1.8927	173.231	>	7.3260986744E-10	***	0.005
III	stand	for	21	1.7508	211.6357	>	1.455407959E-11	***	0.005
III	talk	for	0	17.981	17.981	<	0.000478797289204	***	0.005
III	talk	like	0	18.9274	18.9274	<	0.000184991345369	***	0.005
III	go	to	25	4.9477	81.269	>	3.95570709202015E-06	***	0.005
III	get	from	17	3.8519	44.8798	>	0.02158113936742	*	0.004
III	interested	in	14	0.4723	387.4628	>	6.39354912E-12	***	0.004
III	go	through	15	0.5497	379.8639	>	1.81828742E-12	***	0.004
III	vote	for	10	0.7571	112.84	>	0.000273905494743	***	0.003
III	believe	in	12	0.3837	351.6769	>	4.7381013713E-10	***	0.003
III	feel	like	12	1.0958	108.5066	>	7.25451654066096E-05	***	0.003
III	aware	of	11	0.1273	928.6379	>	1.0080564E-13	***	0.003
III	drive	at	6	0.1321	260.6529	>	0.000211578691654	***	0.002
III	fight	for	8	0.3785	153.467	>	0.000239366806404	***	0.002
III	wait	for	8	0.7571	69.2902	>	0.043868848166452	*	0.002
III	consist	of	6	0.1167	296.6	>	0.000101891135234	***	0.002
III	lead	to	7	0.4981	84.8719	>	0.031346622649406	*	0.002
III	deal	with	9	0.4732	153.6482	>	6.8503365009625E-05	***	0.002
III	make	of	5	0.1379	171.4287	>	0.011911038337238	*	0.001

*A.17: HCFA results for VARIETY * V * P in free RCs/interrogatives (only significant results are shown)*

VARIETY	V	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
CanE	talk	about	38	8.903	95.0955	>	4.687397678001E-08	***	0.008
CanE	come	from	25	3.1961	148.7469	>	1.50482592533E-09	***	0.006
CanE	look	like	20	4.7961	48.1972	>	0.02278753573188	*	0.004
CanE	sound	like	7	0.3027	148.1791	>	0.004561215367488	**	0.002
CanE	think	of	7	0.19	244.0847	>	0.000193152230039	***	0.002
CanE	deal	with	6	0.1514	225.9321	>	0.001888231407427	**	0.002
KenE	come	from	20	1.7711	187.6194	>	8.714242497E-10	***	0.005
KenE	talk	about	20	4.9336	46.0103	>	0.035261941801226	*	0.004
KenE	refer	to	13	0.2767	585.0465	>	8.7506441E-13	***	0.003
BrE	be	like	50	15.748	74.4983	>	5.4953214918127E-07	***	0.009
BrE	talk	about	29	8.4769	49.6877	>	0.003309921217712	**	0.006
BrE	come	from	17	3.0431	64.012	>	0.003337871785621	**	0.004
BrE	think	of	10	0.1809	532.9725	>	1.12206929272E-09	***	0.003
HKE	talk	about	30	4.7318	134.9343	>	8.2815970495E-10	***	0.007
HKE	look	like	16	2.5491	70.9767	>	0.00175912931945	**	0.004
HKE	interested	in	6	0.0803	436.3991	>	4.47013407898437E-05	***	0.002
HKE	go	to	9	0.8412	79.1322	>	0.035031442602059	*	0.002
IndE	consist	of	3	0.0123	725.7196	>	0.039810335228611	*	0.001
IrE	be	like	51	16.9561	68.3522	>	2.25780265860217E-06	***	0.009
IrE	talk	about	30	9.1272	47.7336	>	0.004547250993103	**	0.006
IrE	come	from	22	3.2766	106.9907	>	1.05451248655935E-06	***	0.005
IrE	focus	on	4	0.0551	282.4362	>	0.04722877057341	*	0.001
JamE	be	like	62	18.2893	104.4668	>	9.157963689E-11	***	0.012
JamE	come	from	37	3.5342	316.892	>	5.03409702566141E-20	***	0.009
JamE	talk	about	34	9.8449	59.2661	>	0.000174730048124	***	0.007
JamE	look	at	16	1.758	115.378	>	9.6205093036785E-06	***	0.004
JamE	go	to	12	1.7502	60.0265	>	0.044061058676236	*	0.003
JamE	call	from	8	0.1987	306.2923	>	6.48009998223359E-06	***	0.002
JamE	refer	to	8	0.5521	100.4731	>	0.016835166900301	*	0.002
JamE	deal	with	6	0.1674	203.2212	>	0.003405266740172	**	0.002
NigE	talk	about	40	6.7277	164.5504	>	2.5442748E-13	***	0.009
NigE	come	from	23	2.4152	175.4447	>	3.0158006079E-10	***	0.006
NigE	look	like	18	3.6243	57.0209	>	0.007513930795738	**	0.004
NigE	look	at	11	1.2014	79.9172	>	0.008032724610732	**	0.003
NigE	feel	like	6	0.2649	124.1652	>	0.04921943300275	*	0.002
NigE	go	through	6	0.1329	259.0133	>	0.000878374717331	***	0.002

A.17 (cont.)

VARIETY	V	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
NigE	refer	to	8	0.3773	154.0036	>	0.000934341168815	***	0.002
NZE	come	from	37	3.7677	293.1193	>	4.288091923734E-19	***	0.009
NZE	talk	about	35	10.4952	57.2152	>	0.000243491926377	***	0.007
NZE	believe	in	6	0.1447	236.9353	>	0.001449178482519	**	0.002
NZE	interested	in	7	0.1781	261.3044	>	0.000123965130286	***	0.002
PhilE	talk	about	50	7.1089	258.7807	>	1.22611948065102E-20	***	0.012
PhilE	come	from	20	2.5521	119.2858	>	6.2110519180421E-07	***	0.005
PhilE	aware	of	10	0.0325	3056.9556	>	4.50863840586259E-17	***	0.003
PhilE	stand	for	9	0.4472	163.5742	>	0.000168930268225	***	0.002
PhilE	go	through	6	0.1404	244.5507	>	0.001214812368875	**	0.002
PhilE	believe	in	5	0.098	245.2	>	0.008948307680992	**	0.001
SgE	talk	about	39	7.5799	130.2422	>	7.077466386E-11	***	0.009
SgE	think	of	6	0.1618	210.6587	>	0.002789719845568	**	0.002
SgE	lead	to	5	0.1357	174.3656	>	0.044049299713193	*	0.001
SgE	fall	under	2	0.0005	7996.0005	>	0.016289184492707	*	0.001

A.18: Absolute frequencies of free RCs/interrogatives included in the HCFA VARIETY_TYPE/VARIETY * V * P

VARIETY	Abs. freq.
NZE	468
JamE	439
IrE	407
CanE	397
BrE	378
SgE	338
PhilE	317
NigE	300
KenE	220
HKE	211
IndE	131
TanE	62

*A.19: Absolute frequencies of finite RCs included in the HCFA VARIETY_TYPE/VARIETY * V * P*

VARIETY	Abs. freq.
JamE	81
NZE	76
BrE	66
SgE	63
NigE	58
PhilE	54
IrE	52
KenE	44
CanE	37
HKE	35
TanE	28
IndE	28

*A.20: HCFA results for VARIETY * V * P in finite RCs (only significant results are shown)*

VARIETY	V	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
NigE	come	from	7	0.2256	203.4242	>	0.000338569350873	***	0.011
HKE	rely	on	7	0.0537	898.5304	>	1.707473925621E-08	***	0.011
BrE	move	towards	2	0.0007	5710.2864	>	0.016744613585542	*	0.003

*A.21: Absolute frequencies of interrogatives included in the HCFA
VARIETY_TYPE/VARIETY * FILLER * P*

VARIETY	Abs. freq.
IrE	289
JamE	281
NZE	280
CanE	230
SgE	229
BrE	224
PhilE	208
NigE	161
HKE	142
KenE	117
IndE	86
TanE	36

A.22: HCFA results for VARIETY_TYPE * FILLER * P in interrogatives (only significant results are shown)

VARIETY_TYPE	FILLER	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
V	where	from	117	24.19	356.085	>	8.70568235557792E-40	***	0.041
V	what	from	10	98.268	79.2856	<	7.21956839420156E-28	***	0.04
V	what	like	202	120.9225	54.3618	>	1.77386685163E-09	***	0.037
V	how	for	46	7.6011	193.9819	>	2.0219285604745E-18	***	0.017
V	where	about	1	31.1429	29.175	<	6.0484726058E-10	***	0.013
V	where	like	0	29.7668	29.7668	<	7.509474444E-11	***	0.013
V	what	to	33	62.6679	14.0452	<	0.017656144295984	*	0.013
V	where	for	0	23.9728	23.9728	<	2.628077503618E-08	***	0.011
V	who	to	27	5.6857	79.9021	>	6.516057022329E-08	***	0.009
V	which	like	0	10.2449	10.2449	<	0.026114667372513	*	0.005
V	who	with	15	2.6426	57.786	>	0.000102846469409	***	0.005
V	how	about	0	9.8746	9.8746	<	0.037732014289897	*	0.004
V	which	on	10	1.9941	32.142	>	0.033661505282229	*	0.004
V	when	by	2	0.0082	483.8131	>	0.025434140848724	*	0.001

VARIETY_TYPE	FILLER	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
IV	where	from	74	12.0596	318.1377	>	7.2966287654311E-31	***	0.027
IV	what	from	4	48.9899	41.3165	<	6.812256E-14	***	0.02
IV	where	about	0	15.5258	15.5258	<	0.000130495677872	***	0.007
IV	where	like	0	14.8397	14.8397	<	0.000259989296046	***	0.007
IV	which	to	18	2.6469	89.0542	>	3.8011410540909E-07	***	0.007
IV	where	for	0	11.9512	11.9512	<	0.00472644070043	**	0.005
IV	why	for	5	0.1619	144.5782	>	0.000612347474026	***	0.002
III	what	about	17 0	92.7512	64.3375	>	8.44309816E-11	***	0.035
III	where	from	97	17.7346	354.2794	>	6.68717256696077E-37	***	0.035
III	what	from	5	72.044	62.391	<	2.53290210901346E-22	***	0.03
III	what	to	17	45.9442	18.2344	<	0.000554725053862	***	0.013
III	where	about	1	22.832	20.8758	<	1.98670354133749E-06	***	0.01

VARIETY_TYPE	FILLER	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
III	how	like	29	6.9195	70.4601	>	2.3002511732751E-07	***	0.01
III	where	like	0	21.8232	21.8232	<	2.2959464011565E-07	***	0.01
III	what	of	35	15.3147	25.3032	>	0.007757138105966	**	0.009
III	where	for	2	17.5753	13.8029	<	0.002890376638928	**	0.007
III	where	to	27	11.3098	21.7672	>	0.036742530148279	*	0.007
III	who	with	16	1.9374	102.0733	>	2.2533288828886E-07	***	0.006

A.23: HCFA results for VARIETY * FILLER * P in interrogatives (only significant results are shown)

VARIETY	FILLER	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
CanE	where	from	28	5.4386	93.5933	>	1.921481866772E-08	***	0.01
CanE	what	from	2	22.0935	18.2745	<	0.000192713453943	***	0.009
KenE	where	from	17	2.7666	73.2269	>	2.02554558910377E-05	***	0.006
BrE	what	from	2	21.5171	17.703	<	0.000327612701463	***	0.009
BrE	how	for	11	1.6644	52.3633	>	0.004545767160274	**	0.004
HKE	what	from	0	13.6403	13.6403	<	0.003546643088091	**	0.006
IndE	where	from	12	2.0336	48.844	>	0.0049093272937	**	0.004
IrE	what	from	4	27.7609	20.3373	<	6.91115095417431E-05	***	0.011
IrE	where	from	32	6.8337	92.6793	>	7.17626891911E-09	***	0.011
IrE	what	at	22	6.7324	34.6236	>	0.007284856052334	**	0.007
IrE	how	for	16	2.1473	89.3668	>	3.90283577315354E-06	***	0.006
IrE	when	by	2	0.0023	1735.1327	>	0.008381426731692	**	0.001
JamE	where	from	55	6.6446	351.9015	>	3.75492025847183E-28	***	0.021
JamE	what	like	60	33.2153	21.5991	>	0.048791111914421	*	0.012
JamE	what	from	3	26.9925	21.3259	<	1.90229618090548E-05	***	0.011
NigE	where	from	28	3.807	153.7434	>	4.16498861E-12	***	0.011

VARIETY	FILLER	P	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
NigE	what	from	0	15.4654	15.4654	<	0.00056542588885	***	0.007
NZE	where	from	41	6.6209	178.5139	>	5.19262773658981E-16	***	0.015
NZE	what	from	2	26.8964	23.0451	<	2.20477987600641E-06	***	0.011
NZE	how	for	15	2.0805	80.2276	>	1.95019869670797E-05	***	0.006
Phile	what	from	2	19.9802	16.1804	<	0.001339122681844	**	0.008
Phile	where	from	23	4.9184	66.4737	>	8.41461931892913E-06	***	0.008
Phile	who	with	11	0.5373	203.7374	>	5.032462797945E-08	***	0.005
SgE	what	from	1	21.9974	20.0429	<	1.81518549536395E-05	***	0.009
SgE	where	from	19	5.415	34.0817	>	0.012884558385276	*	0.006
SgE	which	to	9	1.1885	51.3416	>	0.013769037861441	*	0.003
SgE	why	for	5	0.0727	333.9517	>	4.933197505523E-05	***	0.002

*A.24: Absolute frequencies of finite RCs included in the HCFA VARIETY_TYPE/VARIETY * WH * P*

VARIETY	Abs. freq.
BrE	688
IndE	609
JamE	561
IrE	423
KenE	385
NZE	358
NigE	330
TanE	328
PhilE	319
CanE	303
HKE	246
SgE	245

A.25: Sample size per VARIETY(_TYPE) for LNRE models of the V + P slot

VARIETY(_TYPE)	Abs. freq.
V	1,887
III	1,489
IV	923
GB	445

A.26: Sample size per VARIETY(_TYPE) for LNRE models of the V + P slot in informal/medium texts about everyday topics

VARIETY(_TYPE)	Abs. freq.
V	1,029
III	410
IV	324
GB	220

A.27: Sample size per VARIETY(_TYPE) for LNRE models of the FILLER + P slot in interrogatives

VARIETY(_TYPE)	Abs. freq.
V	1,023
III	750
IV	510
GB	224

A.28: Sample size per VARIETY(_TYPE) for LNRE models of the FILLER + P slot in free RCs

VARIETY(_TYPE)	Abs. freq.
V	646
III	499
IV	274
GB	158

A.29: Sample size per VARIETY(_TYPE) for LNRE models of the FILLER + P slot in finite RCs

VARIETY(_TYPE)	Abs. freq.
V	2,217
III	1,772
IV	806
GB	688

A.30: HCFA results for VARIETY * NONSTANDARD_TYPE * CLAUSE_TYPE * PP_TYPE * CATEGORY (only significant results are shown)

VARIETY	NONSTANDARD_TYPE	CLAUSE_TYPE	PP_TYPE	CATEGORY	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
HKE	.	.	.	.	82	58.875	9.0831	>	0.008859014494957	**	0.056
JamE	.	.	.	.	36	58.875	8.8877	<	0.003817678867438	**	0.056
IndE	.	.	.	.	79	58.875	6.8792	>	0.02512640814245	*	0.049
KenE	.	.	.	.	42	58.875	4.8368	<	0.044884127522504	*	0.041
.	missing	.	.	.	260	117.75	171.8477	>	1.21337031702316E-43	***	0.403
.	resumptive	.	.	.	27	117.75	69.9411	<	4.36341003909697E-28	***	0.257
.	doubling	.	.	.	61	117.75	27.3508	<	1.5330270557E-10	***	0.161
.	.	finite RC	.	.	296	94.2	432.3062	>	2.83486495494475E-90	***	0.536
.	.	cleft RC	.	.	2	94.2	90.2425	<	6.37976106820648E-42	***	0.245
.	.	embedded interrogative	.	.	29	94.2	45.1278	<	4.91599189834034E-17	***	0.173
.	.	free RC	.	.	66	94.2	8.442	<	0.000915545255611	***	0.075
.	.	main clause interrogative	.	.	78	94.2	2.786	<	0.033075805202261	*	0.043

A.30 (cont.)

VARIETY	NONSTANDARD_ TYPE	CLAUSE_TYPE	PP_TYPE	CATEGORY	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
.	.	.	optional complement	.	196	26.1667	1102.2922	>	7.90879658118903E-115	***	0.382
.	.	.	prepositional X	.	82	26.1667	119.1345	>	1.08571582840914E-18	***	0.126
.	.	.	respect	.	60	26.1667	43.7461	>	4.253823290095E-08	***	0.076
.	.	.	direction	.	1	26.1667	24.2049	<	9.336475218E-10	***	0.057
.	.	.	accompaniment	.	3	26.1667	20.5106	<	1.0395379215659E-07	***	0.052
.	.	.	affected location	.	3	26.1667	20.5106	<	1.0395379215659E-07	***	0.052
.	.	.	degree	.	3	26.1667	20.5106	<	1.0395379215659E-07	***	0.052
.	.	.	duration/frequency	.	4	26.1667	18.7782	<	6.0134537270586E-07	***	0.05
.	.	.	subcategorized PP	.	4	26.1667	18.7782	<	6.0134537270586E-07	***	0.05
.	.	.	idiosyncratic	.	7	26.1667	14.0393	<	5.43809635651996E-05	***	0.043
.	.	.	time	.	7	26.1667	14.0393	<	5.43809635651996E-05	***	0.043
.	.	.	cause/reason /purpose/result	.	10	26.1667	9.9884	<	0.001354527555357	**	0.036
.	.	.	V-X-P	.	10	26.1667	9.9884	<	0.001354527555357	**	0.036
.	.	.	means/instrument	.	13	26.1667	6.6253	<	0.014310484388118	*	0.03

A.30 (cont.)

VARIETY	NONSTANDARD_ TYPE	CLAUSE_TYPE	PP_TYPE	CATEGORY	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
.	.	.	.	less_spontaneous	162	235.5	22.9395	<	5.92936942E-12	***	0.312
.	.	.	.	more_spontaneous	309	235.5	22.9395	>	1.185873884E-11	***	0.312
.	doubling	.	prepositional X	.	28	10.62	28.443	>	0.000351512625755	***	0.038
.	.	main clause interrogative	respect	.	0	9.9363	9.9363	<	0.003915536455407	**	0.022
.	.	finite RC	.	more_spontaneous	164	194.1911	4.6938	<	0.020341900411243	*	0.109
.	.	finite RC	.	less_spontaneous	132	101.8089	8.9531	>	0.005551495854893	**	0.082
.	.	main clause interrogative	.	less_spontaneous	10	26.828	10.5554	<	0.001329709590287	**	0.038
IndE	.	main clause interrogative	.	more_spontaneous	20	8.583	15.1868	>	0.041963180527306	*	0.025
.	missing	finite RC	prepositional X	.	8	28.447	14.6968	<	0.001443468301165	**	0.046
.	missing	free RC	optional complement	.	35	15.1611	25.96	>	0.002124194991664	**	0.044

A.30 (cont.)

VARIETY	NONSTANDARD_ TYPE	CLAUSE_TYPE	PP_TYPE	CATEGORY	Frequency	Expected	Cont.chisq	Obs.exp	P.adj.Holm	Significance	Q
.	unexpected	finite RC	.	less_spontaneous	51	26.587	22.4168	>	0.000339306162274	***	0.055
.	missing	main clause interrogative	.	more_spontaneous	50	28.2478	16.7503	>	0.003107647438702	**	0.049
.	doubling	.	prepositional X	less_spontaneous	19	3.6527	64.4837	>	1.44001812548754E-06	***	0.033
IndE	missing	main clause interrogative	.	more_spontaneous	17	4.738	31.7342	>	0.002784723831293	**	0.026
.	missing	free RC	optional complement	more_spontaneous	30	9.9465	40.4306	>	0.000105210167782	***	0.043
.	doubling	finite RC	prepositional X	less_spontaneous	14	2.2956	59.6763	>	9.6451080907508E-05	***	0.025

A.31: Top 5 prepositions per ICE variety (absolute and relative frequencies)

VARIETY	PREPOSITION	Abs. freq.	Rel. freq.
BrE	of	29,315	26.51%
	in	18,886	17.08%
	to	9,875	8.93%
	for	8,238	7.45%
	on	6,539	5.91%
	other	37,740	34.12%
CanE	of	24,991	24.50%
	in	17,244	16.91%
	to	9,250	9.07%
	for	8,286	8.12%
	on	6,178	6.06%
	other	36,047	35.34%
IrE	of	26,653	24.99%
	in	18,960	17.77%
	to	9,837	9.22%
	for	8,142	7.63%
	on	6,531	6.12%
	other	36,546	34.26%
NZE	of	26,718	23.91%
	in	18,335	16.41%
	to	10,684	9.56%
	for	9,132	8.17%
	on	6,883	6.16%
	other	39,985	35.78%
JamE	of	27,370	25.65%
	in	18,052	16.92%
	to	10,856	10.17%
	for	8,482	7.95%
	on	6,077	5.70%
	other	35,871	33.62%
SgE	of	23,879	24.11%
	in	16,186	16.34%
	to	9,459	9.55%
	for	9,295	9.38%
	on	5,383	5.43%
	other	34,851	35.18%
KenE	of	22,537	25.54%
	in	15,833	17.94%
	to	8,684	9.84%
	for	7,285	8.25%
	on	4,847	5.49%
	other	29,067	32.94%

A.31 (cont.)

VARIETY	PREPOSITION	Abs. freq.	Rel. freq.
TanE	of	22,382	29.16%
	in	14,465	18.84%
	to	6,629	8.64%
	for	6,063	7.90%
	on	3,819	4.97%
	other	23,408	30.49%
HKE	of	27,889	23.88%
	in	21,600	18.49%
	to	11,515	9.86%
	for	10,834	9.28%
	on	6,369	5.45%
	other	38,582	33.04%
IndE	of	31,912	27.50%
	in	22,028	18.98%
	to	10,993	9.47%
	for	8,632	7.44%
	on	5,753	4.96%
	other	36,723	31.65%
NigE	of	27,876	27.03%
	in	17,790	17.25%
	to	10,066	9.76%
	for	9,137	8.86%
	on	5,744	5.57%
	other	32,530	31.54%
PhilE	of	28,851	26.47%
	in	18,574	17.04%
	to	10,301	9.45%
	for	9,356	8.58%
	with	5,904	5.42%
	other	36,003	33.03%

A.32: PREPOSITION_PLACEMENT in extra tokens (absolute and relative frequencies)

PREPOSITION_PLACEMENT	Abs. freq.	Rel. freq.
pied-piping	54	70.13%
stranding	21	27.27%
mismatching	1	1.30%
resumptive	1	1.30%
Total	77	100%

References

- Adejare, Roseline. 2021. Preposition pied piping and stranding in academic and popular Nigerian English writing. *International Journal of English Linguistics* 11(4). 40–57. <https://doi.org/10.5539/ijel.v11n4p40>.
- Adesola, Oluseye. 2006. Yoruba: A grammar sketch: Version 1.0. <https://afranaphproject.afranaphdatabase.com/images/stories/downloads/casefiles/YorubaGS.pdf>. (8 August, 2023).
- Aitchison, Jean. 2001. *Language change: Progress or decay?* 3rd edn. Cambridge: Cambridge University Press.
- Allen, Cynthia. 1980. Movement and deletion in Old English. *Linguistic Inquiry* 11(2). 261–323.
- Allen, Kachina, Francisco Pereira, Matthew Botvinick & Adele E. Goldberg. 2012. Distinguishing grammatical constructions with fMRI pattern analysis. *Brain and Language* 123(3). 174–182. <https://doi.org/10.1016/j.bandl.2012.08.005>.
- Almahammed, Yazan Shaker, Tun Nur Afizah Zainal Ariff & Harison Hanisa Mohd Sidek. 2015. The acquisition of preposition stranding and pied piping in interrogatives by Arab Jordanian EFL speakers. *International Journal of Linguistics* 7(4). 16–26. <https://doi.org/10.5296/ijl.v7i4.8200>.
- Amaechi, Mary & Bolanle E. Arokoyo. 2022. Yoruba *ní* and *sí*: An asymmetry in the class of prepositions. *Language in Africa* 3(3). 87–116. <https://doi.org/doi:10.37892/2686-8946-2022-3-3-87-116>.
- Amaechi, Mary Chimaobi. 2020. *A'-movement dependencies and their reflexes in Igbo*. Universität Potsdam, PhD diss. https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/47152/file/amaechi_diss.pdf. (30 June, 2021).
- Amaechi, Mary & Doreen Georgi. 2019. Quirks of subject (non-)extraction in Igbo. *Glossa: a journal of general linguistics* 4(1). Art. 69. <https://doi.org/10.5334/gjgl.607>.
- Angitso, Michael Terhemmen. 2014. Question-word movement in Tiv. *Language. Text. Society* 4(1). 41–59.
- Appleby, L. L. 1961. *A first Luyia grammar: With exercises*. Dar es Salaam: East African Literature Bureau.
- Aravind, Athulya & Loes Koring. 2023. Experiencer troubles: A reappraisal of the predicate-based asymmetry in child passives. *Language Acquisition* 30(1). 76–100. <https://doi.org/10.1080/10489223.2022.2115373>.

- Arden, Albert H. 1975. *A progressive grammar of the Telugu language*. Madras: The Christian Literature Society.
- Asher, R. E. 1982. *Tamil*. Amsterdam: North-Holland Publishing Company.
- Asher, R. E. & T.C. Kumari. 1997. *Malayalam*. London: Routledge.
- Audring, Jenny. 2019. Mothers or sisters? The encoding of morphological knowledge. *Word Structure*. Edinburgh University Press 12(3). 274–296.
<https://doi.org/10.3366/word.2019.0150>.
- Awobuluyi, Qladele. 1982. *Essentials of Yoruba grammar*. Ibadan: University Press Limited.
- Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bailey, Charles-James N. 1986. Where English cannot put a preposition before a relative or interrogative pronoun. In Gerhard Leitner (ed.), *The English reference grammar: Language and linguistics, writers and readers*, 156–177. Tübingen: Max Niemeyer.
- Balasubramanian, Chandrika. 2009. *Register variation in Indian English*. Amsterdam: John Benjamins.
- Baldie, Brian J. 1976. The acquisition of the passive voice. *Journal of Child Language* 3(3). 331–348. <https://doi.org/10.1017/S0305000900007224>.
- Bao, Zhiming & Huaqing Hong. 2006. Diglossia and register variation in Singapore English. *World Englishes* 25(1). 105–114. <https://doi.org/10.1111/j.0083-2919.2006.00449.x>.
- Barðal, Johanna. 2008. *Productivity: Evidence from case and argument structure in Icelandic*. Amsterdam: John Benjamins.
- Barðal, Johanna. 2011. Lexical vs. structural case: a false dichotomy. *Morphology* 21(3). 619–654. <https://doi.org/10.1007/s11525-010-9174-1>.
- Bardovi-Harlig, Kathleen. 1987. Markedness and salience in second-language acquisition. *Language Learning* 37. 385–407.
- Baroni, Marco & Stefan Evert. 2014. The zipfR package for lexical statistics: A tutorial introduction. <https://zipfr.r-forge.r-project.org/materials/zipfr-tutorial.pdf>. (12 January, 2022).
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.
<https://doi.org/10.1016/j.jml.2012.11.001>.

- Barrett-Keach, Camillia. 1985. *The syntax and interpretation of the relative clause construction in Swahili*. New York: Garland Publishing.
- Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015. Parsimonious mixed models. Arxiv preprint. <https://doi.org/10.48550/arXiv.1506.04967>.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. <https://doi.org/doi:10.18637/jss.v067.i01>.
- Bautista, Lourdes S. 2004. An overview of the Philippine component of the International Corpus of English (ICE-PHI). *Asian Englishes* 7(2). 8–26. <https://doi.org/10.1080/13488678.2004.10801139>.
- Béchet, Christophe. 2020. An empirical perspective on the contact between English and French: A case study on substitutive complex prepositions. *Linguistics Vanguard* 6(s2). 20180051. <https://doi.org/10.1515/lingvan-2018-0051>.
- Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.
- Belsley, David A., Edwin Kuh & Roy E. Welsch. 1980. *Regression diagnostics: Identifying influential data and sources of collinearity*. Hoboken, NJ: John Wiley & Sons Inc. <https://doi.org/10.1002/0471725153>.
- Bencini, Giulia M. L. 2013. Psycholinguistics. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar*, 379–396. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195396683.013.0021>.
- Bennett, Tina L. 1977. Interrogatives. In Ernest R. Byarushengo, Alessandro Duranti & Larry M. Hyman (eds.), *Haya grammatical structure: phonology, grammar, discourse*, 171–188. Los Angeles: University of Southern California. <https://opac.ku.de/s/uei/de/6/0/BV003324695>.
- Benton, Richard A. 1971. *Pangasinan reference grammar*. Honolulu: University of Hawaii Press. <https://opac.ku.de/s/uei/de/6/0/BV007244010>.
- Bergh, Gunnar & Aimo Seppänen. 2000. Preposition stranding with wh-relatives: A historical survey. *English Language and Linguistics* 4(2). 295–316. <https://doi.org/10.1017/S1360674300000265>.
- Bernaisch, Tobias. 2015. *The lexis and lexicogrammar of Sri Lankan English*. Amsterdam: John Benjamins.
- Bernaisch, Tobias. 2022. Comparing generalised linear mixed-effects models,

- generalised linear mixed-effects model trees and random forests: Filled and unfilled pauses in varieties of English. In Julia Schlüter & Ole Schützler (eds.), *Data and methods in corpus linguistics: Comparative approaches*, 163–193. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/9781108589314.007>.
- Bernaisch, 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. <https://doi.org/10.1075/eww.35.1.02ber>.
- Berndt, Rita Sloan, Charlotte C. Mitchum, Martha W. Burton & Anne N. Haendiges. 2004. Comprehension of reversible sentences in aphasia: The effects of verb meaning. *Cognitive Neuropsychology* 21(2). 229–244.
<https://doi.org/10.1080/02643290342000456>.
- Bhatia, Tej K. 1993. *Punjabi: A cognitive-descriptive grammar*. London: Routledge.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511621024.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *The Longman grammar of spoken and written English*. London: Longman.
- Blei, David M. & John D. Lafferty. 2009. Topic models. In Ashok N. Srivastava & Mehran Sahami (eds.), *Text mining: classification, clustering, and applications*, 71–94. Boca Raton, FL: Chapman & Hall/CRC Presss.
- Blumenthal-Dramé, Alice. 2012. *Entrenchment in usage-based theories: What corpus data do and do not reveal about the mind*. Berlin: De Gruyter.
<https://doi.org/10.1515/9783110294002>.
- Bohmann, Axel. 2019. *Variation in English worldwide: registers and global varieties*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/9781108751339>.
- Bolinger, Dwight. 1977. Transitivity and spatiality: The passive of prepositional verbs. In Adam Makkai, Valerie Becker Mackai & Luigi Heilmann (eds.), *Linguistics at the crossroads*, 57–78. Padova: Liviana Editrice.
- Bolt, Philip & Kingsley Bolton. 1996. The International Corpus of English in Hong Kong. In Sidney Greenbaum (ed.), *Comparing English worldwide: The International Corpus of English*, 197–214. Oxford: Clarendon Press.
<https://doi.org/10.1093/oso/9780198235828.003.0014>.
- Bolton, Kingsley. 2012. Language policy and planning in Hong Kong: the historical context and current realities. In Ee Ling Low & Azirah Hashim (eds.), *English in*

- Southeast Asia: features, policy and language in use*, 221–238. Amsterdam: John Benjamins.
- Braine, Martin D. S. & Melissa Bowerman. 1976. Children's first word combinations. *Monographs of the Society for Research in Child Development* 41(1). 1–104. <https://doi.org/10.2307/1165959>.
- Brauer, Markus & John J. Curtin. 2018. Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods* 23(3). 389–411. <https://doi.org/10.1037/met0000159>.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene M. Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.
- Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 168–213.
- Brinton, Laurel J. & Elizabeth Closs Traugott. 2005. *Lexicalization and language change*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511615962>.
- Brown, Keith & Jim Miller. 1991. *Syntax: A linguistic introduction to sentence structure*. 2nd edn. London: Routledge.
- Bruckmaier, Elisabeth. 2017. *Getting at GET in World Englishes*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110497311>.
- Brunner, Thomas. 2017. *Simplicity and typological effects in the emergence of New Englishes: The noun phrase in Singaporean and Kenyan English*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110521801>.
- Brunner, Thomas. 2021. Into-causatives in World Englishes. *English World-Wide* 43(1). 1–32. <https://doi.org/10.1075/eww.21068.bru>.
- Brunner, Thomas & Thomas Hoffmann. 2020. The way-construction in World Englishes. *English World-Wide* 41(1). 1–32.
- Budohoska, Natalia. 2014. *English in Kenya or Kenyan English?* Frankfurt am Main: Peter Lang.
- Bujang, Mohamad Adam & Nurakmal Baharum. 2017. Guidelines of the minimum sample size requirements for Kappa agreement test. *Epidemiology, Biostatistics, and Public Health* 14(2). <https://doi.org/10.2427/12267>.
- Buregeya, Alfred. 2019. *Kenyan English*. Berlin: De Gruyter.

- <https://doi.org/10.1515/9781614516255>.
- Buschfeld, Sarah, Thomas Hoffmann, Magnus Huber & Alexander Kautzsch. 2014. Introduction. In Sarah Buschfeld, Thomas Hoffmann, Magnus Huber & Alexander Kautzsch (eds.), *The evolution of Englishes: The Dynamic Model and beyond*, 1–18. Amsterdam: John Benjamins.
<https://doi.org/10.1075/veaw.g49.01int>.
- Buschfeld, Sarah & Alexander Kautzsch. 2020. Theoretical models of English as a world language. In Daniel Schreier, Edgar W. Schneider & Marianne Hundt (eds.), *The Cambridge handbook of World Englishes*, 51–71. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108349406.003>.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511750526>.
- Cacoullos, Rena Torres & James A. Walker. 2009. On the persistence of grammar in discourse formulas: A variationist study of that. 47(1). 1–43.
<https://doi.org/10.1515/LING.2009.001>.
- Callies, Marcus. 2018. Towards a process-oriented approach to comparing EFL and ESL varieties: A corpus-study of lexical innovations. In Sandra C. Deshors, Sandra Götz & Samantha Laporte (eds.), *Rethinking linguistic creativity in non-native Englishes*, 99–120. Amsterdam: John Benjamins.
<https://doi.org/10.1075/bct.98.05cal>.
- Cappelle, Bert. 2006. Particle placement and the case for “allostructions.” *Constructions*. <https://doi.org/10.24338/cons-381>.
- Cappelle, Bert. 2009. Can we factor out free choice? In Andreas Dufter, Jürg Fleischer & Guido Seiler (eds.), *Describing and modeling variation in grammar*, 183–201. Berlin: De Gruyter.
- Cappelle, Bert, Yury Shtyrov & Friedemann Pulvermüller. 2010. Heating up or cooling up the brain? MEG evidence that phrasal verbs are lexical units. *Brain and Language* 115(3). 189–201. <https://doi.org/10.1016/j.bandl.2010.09.004>.
- Census and Statistics Department. 2022. *2021 Population census: summary results*. Hong Kong: Hong Kong Special Administrative Region.
<https://www.census2021.gov.hk/doc/pub/21c-summary-results.pdf>. (1 August, 2022).
- Chan, Alice Y. W. 2004. Syntactic transfer: Evidence from the interlanguage of Hong Kong Chinese ESL learners. *The Modern Language Journal* 88(1). 56–74.
- Chen, Weirong. 2020. *A grammar of Southern Min: The Hui'an dialect*. Berlin: De

Gruyter.

- Cheung, Candice Chi-Hang. 2008. *Wh-fronting in Chinese*. University of Southern California, PhD diss. <https://doi.org/10.13140/RG.2.1.4763.7361>.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: The MIT Press. <https://mitpress.mit.edu/9780262531283/the-minimalist-program/>. (5 January, 2024).
- Clahsen, Harald & Claudia Felser. 2006a. How native-like is non-native language processing? *Trends in Cognitive Sciences* 10. 564–570.
- Clahsen, Harald & Claudia Felser. 2006b. Grammatical processing in language learners. *Applied Psycholinguistics* 27(1). 3–42. <https://doi.org/10.1017/S0142716406060024>.
- Clark, Herbert & Eve Clark. 1977. *Psychology and linguistics: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.
- Colleman, Timothy. 2016. A reflection on constructionalization and constructional borrowing, inspired by an emerging Dutch replica of the ‘time’-away construction. *Belgian Journal of Linguistics* 30(1). 91–113. <https://doi.org/10.1075/bjl.30.05col>.
- Colleman, Timothy. 2018. On contact-related semantic shifts in Afrikaans three-argument constructions: Distributional assimilation in constructional semantics. In Hans C. Boas & Steffen Höder (eds.), *Constructions in contact: Constructional perspectives on contact phenomena in Germanic languages*, 143–177. Amsterdam: John Benjamins. <https://doi.org/10.1075/cal.24.05col>.
- Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgements*. Thousands Oaks: Sage.
- Croft, William. 2000. *Explaining language change*. Harlow: Longman.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198299554.001.0001>.
- Croft, William & D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511803864>.
- Crystal, David. 2003. *English as a global language*. 2nd edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486999>.
- Cunnings, Ian. 2017. Parsing and working memory in bilingual sentence processing.

- Bilingualism: Language and Cognition* 20(4). 659–678.
<https://doi.org/10.1017/S1366728916000675>.
- Dąbrowska, Eva. 2000. From formula to schema: The acquisition of English questions. *Cognitive Linguistics* 11(1–2). 83–102. <https://doi.org/10.1515/cogl.2001.013>.
- Dallas, Andrea & Edith Kaan. 2008. Second language processing of filler-gap dependencies by late learners. *Language and Linguistics Compass* 2(3). 372–388. <https://doi.org/10.1111/j.1749-818X.2008.00056.x>.
- David, Anne Boyle. 2015. *Descriptive grammar of Bangla*. Berlin: De Gruyter.
<https://doi.org/10.1515/9781614512295>.
- 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.
<https://doi.org/10.1017/CBO9780511597855.002>.
- 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*, 102–119. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9781107477186.008>.
- Degenhardt, Julia & Tobias Bernaisch. 2022. Apologies in South Asian varieties of English: A corpus-based study on Indian and Sri Lankan English. *Corpus Pragmatics* 6. 201–223. <https://doi.org/10.1007/s41701-022-00117-8>.
- Department of Statistics Singapore. 2021. *Census of population 2020: statistical release 1. Demographic characteristics, education, language and religion*.
https://www.singstat.gov.sg/publications/reference/cop2020/cop2020-sr1/census20_stat_release1. (10 June, 2023).
- Deshors, Sandra C., Sandra Götz & Samantha Laporte. 2018. Linguistic innovations in EFL and ESL: Rethinking the linguistic creativity of non-native English speakers. In Sandra C. Deshors, Sandra Götz & Samantha Laporte (eds.), *Linguistic innovations in EFL and ESL: rethinking the linguistic creativity of non-native English speakers*, 1–20. Amsterdam: John Benjamins.
<https://doi.org/10.1075/ijlcr.2.2.01des>.
- Deterding, David. 2007. *Singapore English*. Edinburgh: Edinburgh University Press.
- Dhongde, Ramesh Vaman & Kashi Wali. 2009. *Marathi*. Amsterdam: John Benjamins.
- Diessel, Holger. 2004. *The acquisition of complex sentences*. Cambridge: Cambridge

- University Press. <https://doi.org/10.1017/CBO9780511486531>.
- Diessel, Holger. 2013. Construction Grammar and first language acquisition. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar*, 346–364. Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780195396683.013.0019>.
- Diessel, Holger. 2019. *The grammar network: How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/9781108671040>.
- Dimmendaal, Gerrit Jan. 1983. *The Turkana language*. Dordrecht: Foris.
- Dor, Daniel. 2005. Toward a semantic account of *that*-deletion in English. *Linguistics* 43(2). 345–382. <https://doi.org/10.1515/ling.2005.43.2.345>.
- Dubois, Tanguy, Magali Paquot & Benedikt Szmrecsanyi. 2022. Alternation phenomena and language proficiency: The genitive alternation in the spoken language of EFL learners. *Corpus Linguistics and Linguistic Theory*.
<https://doi.org/10.1515/cllt-2021-0078>.
- Durleman-Tame, Stephanie. 2008. *The syntax of Jamaican Creole: A cartographic perspective*. Amsterdam: John Benjamins.
<https://opac.ku.de/s/uei/de/6/0/BV023392947>.
- Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41(1). 87–100. <https://doi.org/10.1146/annurev-anthro-092611-145828>.
- Edwards, Alison. 2016. *English in the Netherlands*. Amsterdam: John Benjamins.
- Edwards, Alison & Samantha Laporte. 2015. Outer and expanding circle Englishes: The competing roles of norm orientation and proficiency levels. *English World-Wide*. John Benjamins 36(2). 135–169. <https://doi.org/10.1075/eww.36.2.01edw>.
- Ejeba, Salem Ochala. 2016. *A grammar of Ígálâ*. Port Harcourt: M & J Grand Orbit Communications Ltd. <https://opac.ku.de/s/uei/de/6/0/BV043992237>.
- Ellis, Nick C. 2003. Constructions, chunking, and connectionism: The emergence of second language structure. In Catherine Doughty & Michael H. Long (eds.), *Handbook of second language acquisition*, 33–68. Oxford: Blackwell Publishing.
- Ellis, Nick C. 2006. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1). 1–24. <https://doi.org/10.1093/applin/ami038>.
- Ellis, Nick C. 2013. Construction Grammar and second language acquisition. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of*

- Construction Grammar*, 365–378. Oxford: Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780195396683.013.0020>.
- Ellis, Rod. 2008. *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, Rod. 2015. *Understanding second language acquisition*. Oxford: Oxford University Press.
- Emenanjo, E. Nolue. 2015. *A grammar of contemporary Igbo: Constituents, features and processes*. Part Harcourt: M & J Grand Orbit Communications Ltd.
- Engelhardt, Paul E., Martin Corley, Joel T. Nigg & Fernanda Ferreira. 2010. The role of inhibition in the production of disfluencies. *Memory & Cognition* 38(5). 617–628. <https://doi.org/10.3758/MC.38.5.617>.
- Ernst, Thomas. 2002. *The syntax of adjuncts*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511486258>.
- Essien, Okon E. 1990. *A grammar of the Ibibio language*. Ibadan: University Press Limited.
- Evert, Stefan & Marco Baroni. 2007. zipfR: Word frequency distributions in R. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Posters and Demonstrations Sessions*, 29–32.
- Evert, Stefanie. 2004. A simple LNRE model for random character sequences. In Purnelle, Cédric Fairon & Anne Dister (eds.), *Proceedings of JADT 2004*, 411–422. Louvain: Presses universitaires de Louvain.
- Fannami, Muhammad. 2011. *An introduction to morphology and syntax of the Kanuri language*. München: Lincom Europa.
- Faraclas, Nicholas. 1996. *Nigerian Pidgin*. London: Routledge.
- Faraclas, Nicholas. 2008. Nigerian Pidgin English: Morphology and syntax. In Rajend Mesthrie, Bernd Kortmann & Edgar W. Schneider (eds.), *Africa, South and Southeast Asia*, vol. 4, 340–367. Berlin: De Gruyter.
<https://doi.org/10.1515/9783110208429.2.340>.
- Feinerer, Ingo & Kurt Hornik. 2020. tm: Text mining package. R package version 0.7-8.
<https://CRAN.R-project.org/package=tm>. (25 January, 2022).
- Felser, Claudia, Ian Cummings, Claire Batterham & Harald Clahsen. 2012. The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition* 34(1). 67–98. <https://doi.org/10.1017/S0272263111000507>.
- Felser, Claudia & Leah Roberts. 2007. Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research* 23(1). 9–

36. <https://doi.org/10.1177/0267658307071600>.
- Ferreira, Fernanda. 2003. The misinterpretation of noncanonical sentences. *Cognitive Psychology* 47(2). 164–203. [https://doi.org/10.1016/S0010-0285\(03\)00005-7](https://doi.org/10.1016/S0010-0285(03)00005-7).
- Fischer, Olga. 1992. Syntax. In Norman Blake (ed.), *The Cambridge history of the English language*, vol. Vol. II: 1066-1476, 207–408. Cambridge: Cambridge University Press.
- Fischer, Olga, Ans van Kemenade, Willem Koopman & Wim van der Wurff. 2000. *The syntax of early English*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511612312>.
- Fox, John. 2003. Effect displays in R for generalised linear models. *Journal of Statistical Software* 8(15). 1–27. <https://doi.org/10.18637/jss.v008.i15>.
- Fox, John & Georges Monette. 1992. Generalized collinearity diagnostics. *Journal of the American Statistical Association* 87(417). 178–183.
<https://doi.org/10.2307/2290467>.
- Fox, John & Sanford Weisberg. 2019. *An R companion to applied regression*. 3rd edn. Thousands Oaks: Sage.
<https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- Funke, Nina & Tobias Bernaisch. 2022. Intensifying and downtoning in South Asian Englishes: Empirical perspectives. *English World-Wide* 43(1). 33–65.
<https://doi.org/10.1075/eww.21064.fun>.
- Gardner, Matt Hunt, Eva Uffing, Nicholas Van Vaecck & Benedikt Szmrecsanyi. 2021. Variation isn't that hard: Morphosyntactic choice does not predict production difficulty. *PLOS ONE* 16(6). e0252602.
<https://doi.org/10.1371/journal.pone.0252602>.
- Gass, Susan M. 1996. Second language acquisition and linguistic theory: The role of language transfer. In William C. Ritchie & Tej K. Bhatia (eds.), *Handbook of second language acquisition*, 317–345. San Diego, CA: Academic Press.
https://doi.org/10.1007/978-94-009-2733-9_21.
- Gass, Susan M., Jennifer Behney & Luke Plonsky. 2020. *Second language acquisition: An introductory course*. New York: Routledge.
<https://doi.org/10.4324/9781315181752>.
- Gelman, Andrew & Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511790942>.
- Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies.

- Cognition* 68(1). 1–76. [https://doi.org/10.1016/S0010-0277\(98\)00034-1](https://doi.org/10.1016/S0010-0277(98)00034-1).
- Gilquin, Gaëtanelle. 2015. The use of phrasal verbs by French-speaking EFL learners: A constructional and collostructional corpus-based approach. *Corpus Linguistics and Linguistic Theory* 11(1). 51–88. <https://doi.org/10.1515/cllt-2014-0005>.
- Gilquin, Gaëtanelle. 2017. Applied cognitive linguistics and second/foreign language varieties: Towards an explanatory account. In Jacqueline Evers-Vermeul & Elena Tribushinina (eds.), *Usage-based approaches to language acquisition and language teaching*, 47–72. Berlin: De Gruyter.
<https://doi.org/10.1515/9781501505492-003>.
- Gilquin, Gaëtanelle & Sylviane Granger. 2011. Evidence from the International Corpus of Learner English: From EFL to ESL. In Joybrato Mukherjee & Marianne Hundt (eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*, 55–78. Amsterdam: John Benjamins.
<https://doi.org/10.1075/scl.44.04gra>.
- Gilquin, Gaëtanelle & Magali Paquot. 2008. Too chatty: learner academic writing and register variation. *English Text Construction* 1(1). 41–61.
<https://doi.org/10.1075/etc.1.1.05gil>.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago, IL: University of Chicago Press.
- Goldberg, Adele E. 2002. Surface generalizations: An alternative to alternations. 13(4). 327–356. <https://doi.org/10.1515/cogl.2002.022>.
- Goldberg, Adele E. 2003. Constructions: A new theoretical approach to language. 7(5). 219–224. [https://doi.org/10.1016/s1364-6613\(03\)00080-9](https://doi.org/10.1016/s1364-6613(03)00080-9).
- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199268511.001.0001>.
- Goldberg, Adele E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.
<https://doi.org/10.1515/9780691183954>.
- Gonzalez, Andrew. 1991. Stylistic shifts in the English of the Philippine print media. In Jenny Cheshire (ed.), 333–363. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511611889.024>.
- Görlach, Manfred. 1988. The development of Standard Englishes. In Manfred Görlach (ed.), *Studies in the history of the English language*, 9–64. Heidelberg: Winter.
- Grafmiller, Jason & Benedikt Szmrecsanyi. 2018. Mapping out particle placement in

- Englishes around the world: A study in comparative sociolinguistic analysis. *Language Variation and Change* 30(3). 385–412.
<https://doi.org/10.1017/S0954394518000170>.
- Green, Margaret M. & G. E. Igwe. 1963. *A descriptive grammar of Igbo*. Berlin: Akademie-Verlag.
- Greenbaum, Sidney & Gerald Nelson. 1996. The International Corpus of English (ICE) Project. *World Englishes* 15(1). 3–15.
- Griefenow-Mewis, Catherine. 2001. *A grammatical sketch of written Oromo*. Köln: Köppe.
- Gries, Stefan Th. 2002. Preposition stranding in English: Predicting speakers' behaviour. In Vida Samiian (ed.), *Proceedings of the Western Conference on Linguistics*, vol. 12, 230–241. Fresno, CA: California State University.
- Gries, Stefan Th. 2003. *Multifactorial analysis in corpus linguistics: A study of particle placement*. London: Continuum Press.
- Gries, Stefan Th. 2004. HCFA 3.2 – A program for Hierarchical Configurational Frequency Analysis.
- Gries, Stefan Th. 2009. *Statistics for linguistics with R: A practical introduction*. Berlin: De Gruyter.
- Gries, Stefan Th. 2021. (Generalized linear) mixed-effects modeling: A learner corpus example. *Language Learning* 71(3). 757–798.
<https://doi.org/10.1111/lang.12448>.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspective on “alternations.” *International Journal of Corpus Linguistics* 9(1). 97–129. <https://doi.org/10.1075/ijcl.9.1.06gri>.
- Gries, Stefan Th & Stefanie Wulff. 2005. Do foreign language learners also have constructions? *Annual Review of Cognitive Linguistics* 3(1). 182–200.
<https://doi.org/10.1075/arcl.3.10gri>.
- Gries, Stefan Th & Stefanie Wulff. 2009. Psycholinguistic and corpus-linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics* 7(1). 163–186. <https://doi.org/10.1075/arcl.7.07gri>.
- Gries, Stefan Th. & Stefanie Wulff. 2013. The genitive alternation in Chinese and German ESL learners: Towards a multifactorial notion of context in learner corpus research. *International Journal of Corpus Linguistics* 18(3). 327–356.
<https://doi.org/10.1075/ijcl.18.3.04gri>.
- Gropen, Jess, Steven Pinker, Michelle Hollander, Richard Goldberg & Ronald Wilson.

1989. The learnability and acquisition of the dative alternation in English. *Language* 65(2). 203–257. <https://doi.org/10.2307/415332>.
- Grün, Bettina & Kurt Hornik. 2011. topicmodels: An R package for fitting topic models. *Journal of Statistical Software* 40(13). 1–30. <https://doi.org/10.18637/jss.v040.i13>.
- Grün, Bettina & Kurt Hornik. 2022. Topicmodels: topic models. <https://CRAN.R-project.org/package=topicmodels>. (25 January, 2022).
- Gundel, Jeanette K. 1988. Universals of topic-comment structure. In Michael Hammond, Edith A. Moravcsik & Jessica Wirth (eds.), *Studies in syntactic typology*, 209–239. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.17.16gun>.
- Gut, Ulrike. 2004. Nigerian English: phonology. In Bernd Kortmann & Edgar W. Schneider (eds.), *A handbook of varieties of English: a multimedia reference tool. Volume 1: phonology. Volume 2: morphology and syntax*, 813–830. Berlin: De Gruyter. <https://doi.org/10.1515/9783110197181-051>.
- Gut, Ulrike. 2011. Studying structural innovations in New English varieties. In Joybrato Mukherjee & Marianne Hundt (eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*, 101–124. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.44.06gut>.
- Gut, Ulrike. 2014. *ICE Nigeria*.
- Gut, Ulrike. 2017. English in West Africa. In Markku Filppula, Juhani Klemola & Devyani Sharma (eds.), *The Oxford handbook of World Englishes*, 491–507. Oxford: Oxford University Press.
- Gut, Ulrike & Lilian Coronel. 2012. Relatives worldwide. In Marianne Hundt & Ulrike Gut (eds.), *Mapping unity and diversity world-wide: corpus-based studies of New Englishes*, 215–242. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g43.09gut>.
- Hahn, Udo. 2005. Information retrieval. In Philipp Strazny (ed.), *Encyclopedia of linguistics*, 531–534. New York: Taylor & Francis.
- Hampe, Beate. 2012. Transitive phrasal verbs in acquisition and use: A view from Construction Grammar. *Language Value* 4(1). 1–32. <https://doi.org/10.6035/LanguageV.2012.4.2>.
- Haspelmath, Martin & Susanne M. Michaelis. 2013. Instrument relative clauses. In Susanne M. Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The atlas of Pidgin and Creole language structures*, 374–377. Oxford:

Oxford University Press.

- Hasselgren, Angela. 1994. Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Review of Applied Linguistic* 4(2). 237–260.
- Hawkins, John A. 1999. Processing complexity and filler-gap dependencies across grammars. *Language* 75(2). 244–285. <https://doi.org/10.2307/417261>.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199252695.001.0001>.
- Heine, Bernd. 1982. *Boni dialects*. Berlin: Dietrich Reimer.
- Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511614132>.
- Heller, Benedikt, Benedikt Szmrecsanyi & Jason Grafmiller. 2017. Stability and fluidity in syntactic variation world-wide: The genitive alternation across varieties of English. *Journal of English Linguistics* 45(1). 3–27. <https://doi.org/10.1177/0075424216685405>.
- Hendery, Rachel. 2012. *Relative clauses in time and space: A case study in the methods of diachronic typology*. Amsterdam: John Benjamins.
- Hilpert, Martin. 2013. *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139004206>.
- Hilpert, Martin. 2019. *Construction Grammar and its application to English*. Edinburgh: Edinburgh University Press.
- Himmelman, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In Karl Alexander Adelaar & Nikolaus P. Himmelman (eds.), *The Austronesian languages of Asia and Madagascar*, 110–181. London: Routledge.
- Hinrichs, Lars & Benedikt Szmrecsanyi. 2007. Recent changes in the function and frequency of Standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics* 11(3). 437–474. <https://doi.org/10.1017/S1360674307002341>.
- Höder, Steffen. 2012. A diasystematic approach to common structures: Multilingual constructions. In Kurt Braunmüller & Christoph Gabriel (eds.), *Multilingual Individuals and multilingual societies*, 241–258. Amsterdam: John Benjamins. <https://doi.org/10.1075/hsm.13.17hod>.

- Höder, Steffen. 2014. Grammatical organisation in bilingual groups: Constructing diasystems. In Tor A. Åfarli & Brit Mæhlum (eds.), *The sociolinguistics of grammar*, 137–152. Amsterdam: John Benjamins.
<https://doi.org/10.1075/slcs.154.07hod>.
- Höder, Steffen. 2018. Background and basic concepts of Diasystematic Construction Grammar: Grammar is community-specific. In Hans C. Boas & Steffen Höder (eds.), *Constructions in contact: Constructional perspectives on contact phenomena in Germanic languages*, 37–70. Amsterdam: John Benjamins.
<https://doi.org/10.1075/cal.24.02hod>.
- 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). 165–195. <https://doi.org/10.1515/CLLT.2006.009>.
- Hoffmann, Thomas. 2007a. ‘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*, 161–184. Berlin: De Gruyter.
<https://doi.org/10.1515/9783110198621.161>.
- Hoffmann, Thomas. 2007b. Complements versus adjuncts: A Construction Grammar account of English prepositional phrases. *Occasional Papers in Language and Linguistics* 3. 92–119.
- 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.
<https://doi.org/10.1017/CBO9780511792519.008>.
- Hoffmann, Thomas. 2014. The role of constructions in the Dynamic Model: The cognitive evolution of Englishes. In Sarah Buschfeld, Thomas Hoffmann, Magnus Huber & Alexander Kautzsch (eds.), *The evolution of Englishes: The Dynamic Model and beyond*, 160–180. John Benjamins.
<https://doi.org/10.1075/veaw.g49.10hof>.
- Hoffmann, Thomas. 2019. *English comparative correlatives: Diachronic and synchronic variation at the lexicon-syntax interface*. Cambridge: Cambridge University Press.
- Hoffmann, Thomas. 2020a. Marginal argument structure constructions: The [V *the*

- Ntaboo-word *out of*]-construction in post-colonial Englishes. *Linguistics Vanguard* 6(1). 20190054. <https://doi.org/doi:10.1515/lingvan-2019-0054>.
- Hoffmann, Thomas. 2020b. What would it take for us to abandon Construction Grammar?: Falsifiability, confirmation bias and the future of the constructionist enterprise. *Belgian Journal of Linguistics* 34. 148–160. <https://doi.org/10.1075/bjl.00042.hof>.
- Hoffmann, Thomas. 2021. *The cognitive foundation of post-colonial Englishes: Construction Grammar as the cognitive theory for the Dynamic Model*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108909730>.
- Hoffmann, Thomas. 2022. *Construction Grammar: The structure of English*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781139004213>.
- Hoffmann, Thomas & Thomas Brunner. 2022. Construction Grammar meets the Dynamic Model. In Aloysius Ngefac, Hans-Georg Wolf & Thomas Hoffmann (eds.), *World Englishes and Creole languages today. Vol. 1: The Schneiderian thinking and beyond & Vol. 2: The Bobdian thinking and beyond*, 25–38. München: Lincom.
- Hoffmann, Thomas & Graeme Trousdale. 2013a. Construction Grammar: Introduction. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar*, 1–12. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195396683.013.0001>.
- Hoffmann, Thomas & Graeme Trousdale (eds.). 2013b. *The Oxford handbook of Construction Grammar*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195396683.001.0001>.
- Hopp, Holger. 2006. Syntactic features and reanalysis in near-native processing. *Second Language Research* 22(3). 369–397.
- Hopp, Holger. 2009. The syntax–discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition* 12(4). 463–483. <https://doi.org/10.1017/S1366728909990253>.
- Hopp, Holger. 2010. Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua* 120(4). 901–931. <https://doi.org/10.1016/j.lingua.2009.06.004>.
- Hopp, Holger. 2013. Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research* 29(1). 33–56.

- Hopp, Holger. 2014. Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition* 21(3). 250–278.
- Hornstein, Norbert & Amy Weinberg. 1981. Case theory and preposition stranding. *Linguistic Inquiry* 12(1). 55–91.
- Horsch, Jakob. 2023. *The comparative correlative construction in World Englishes: A Usage-based Construction Grammar approach*. Katholische Universität Eichstätt-Ingolstadt, PhD diss. <https://doi.org/10.17904/ku.opus-833>.
- Huang, C.-T. James, Y.-H. Audrey Li & Yafei Li. 2009. *The syntax of Chinese*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9781139166935>.
- Huber, Magnus & Kari Dako. 2004. Ghanaian English: Morphology and syntax. In Bernd Kortmann & Edgar W. Schneider (eds.), *A handbook of varieties of English: a multimedia reference tool*, vol. 2, 2046–2057. Berlin: De Gruyter.
<https://doi.org/10.1515/9783110197181-122>.
- Huddleston, Rodney D. 2002a. Non-finite and verbless clauses. In Rodney D. Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge grammar of the English language*, 1171–1272. Cambridge: Cambridge University Press.
- Huddleston, Rodney D. 2002b. The clause: Complements. In Rodney D. Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge grammar of the English language*, 213–322. Cambridge: Cambridge University Press.
- Huddleston, Rodney D. 2002c. Clause type and illocutionary force. In Rodney D. Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge grammar of the English language*, 851–946. Cambridge: Cambridge University Press.
- Huddleston, Rodney D., Geoffrey K. Pullum & Peter Peterson. 2002. Relative constructions and unbounded dependencies. In Rodney D. Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge grammar of the English language*, 1031–1096. Cambridge: Cambridge University Press.
- Hudson-Ettle, Diana M. & Josef Schmied. 1999. *Manual to accompany the East African component of the International Corpus of English*. Chemnitz University of Technology.
- Hulstijn, Jan H. & Elaine Marchena. 1989. Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition* 11(3). 241–255.
- Hundt, Marianne. 2006. *The committee has/have decided...: On concord patterns with collective nouns in inner- and outer-circle varieties of English*. *Journal of English Linguistics* 34(3). 206–232. <https://doi.org/10.1177/0075424206293056>.

- Hundt, Marianne, Jennifer Hay & Elizabeth Gordon. 2004. New Zealand English: morphosyntax. In Bernd Kortmann & Edgar W. Schneider (eds.), *A handbook of varieties of English: a multimedia reference tool*, vol. 2, 1752–1784. Berlin: De Gruyter. <https://doi.org/10.1515/9783110197181-108>.
- Hutchison, John P. 1981. *A reference grammar of the Kanuri language*. Madison: University of Wisconsin.
- ICE. International Corpus of English. <https://www.ice-corpora.uzh.ch/en.html>. (20 April, 2021).
- Jach, Daniel. 2018. A usage-based approach to preposition placement in English as a second language. *Language Learning* 68(1). 271–304. <https://doi.org/10.1111/lang.12277>.
- 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*, 339–374. Amsterdam: John Benjamins. <https://doi.org/10.1075/cal.30.11jac>.
- Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1). 23–62. <https://doi.org/10.1016/j.cogpsych.2010.02.002>.
- Jankowski, Bridget L. & Sali A. Tagliamonte. 2014. On the genitive's trail: Data and method from a sociolinguistic perspective. *English Language and Linguistics* 18(2). 305–329. <https://doi.org/10.1017/S1360674314000045>.
- Jäschke, Katja & Ingo Plag. 2016. The dative alternation in German-English interlanguage. *Studies in Second Language Acquisition* 38(3). 485–521. <https://doi.org/10.1017/S0272263115000261>.
- Johansson, Christine & Christer Geisler. 1998. Pied piping in spoken English. In Antoinette Renouf (ed.), *Explorations in corpus linguistics*, 67–82. Amsterdam: Rodopi.
- Jowitt, David. 2019. *Nigerian English*. Berlin: De Gruyter. <https://doi.org/10.1515/9781501504600>.
- Jurafsky, Dan. 1992. An on-line computational model of human sentence interpretation - AAI. In *AAAI-92 Proceedings*, 302–308. San Jose, CA.
- Kaan, Edith. 2014. Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism* 4(2). 257–282. <https://doi.org/10.1075/lab.4.2.05kaa>.

- Kachru, Braj B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In Randolph Quirk & Henry G. Widdowson (eds.), *English in the world: teaching and learning the language and literatures*, 11–30. Cambridge: Cambridge University Press for The British Council.
- Kachru, Yamuna. 2006. *Hindi*. Amsterdam: John Benjamins.
<https://doi.org/10.1075/loall.12>.
- Kader, Mashudi B. H. 1981. *The syntax of Malay interrogatives*. Kuala Lumpur: Dewan Bahasa dan Pustak.
- Kallen, Jeffrey L. & John M. Kirk. 2008. *ICE-Ireland: A user's guide*. Queen's University Belfast.
- Kaltenböck, Gunther. 2006. '... *That* is the question': Complementizer omission in extraposed *that*-clauses. *English Language & Linguistics* 10(2). 371–396.
<https://doi.org/10.1017/S1360674306001961>.
- Kamwangamalu, Nkonko & Alla Tovaes. 2016. English in language ideologies, attitudes, and educational practices in Kenya and South Africa. *World Englishes* 35(3). 421–439. <https://doi.org/10.1111/weng.12207>.
- Kandybowicz, Jason. 2008. *The grammar of repetition: Nupe grammar at the syntax-phonology interface*. Amsterdam: John Benjamins.
- Kao, Rong-Rong. 2001. Where have the prepositions gone? A study of English prepositional verbs and input enhancement in instructed SLA. *International Review of Applied Linguistics in Language Teaching* 39(3). 195–215.
<https://doi.org/10.1515/iral.2001.002>.
- 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). 333–351. <https://doi.org/10.2307/412588>.
- Kellerman, Eric. 1979. Transfer and non-transfer: Where we are now. *Studies in Second Language Acquisition*. Cambridge University Press 2(1). 37–57.
- Kellerman, Eric. 1983. Now you see it, now you don't. In Susan M. Gass & Larry Selinker (eds.), 112–134. Rowley, MA: Newbury House Publishers.
- Kharkwal, Gaurav & Karin Stromswold. 2014. Good-enough language processing: evidence from sentence-video matching. *Journal of Psycholinguistic Research* 43(1). 27–43. <https://doi.org/10.1007/s10936-013-9239-5>.
- Kirkpatrick, Andy & Wang Lixun. 2021. *Is English an Asian language?* Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316471166>.

- Kleinmann, Howard H. 1977. Avoidance behavior in adult second language acquisition. *Language Learning* 27. 93–107.
- Kluender, Robert. 1990. A neurophysiological investigation of wh-islands. In *Proceedings of the sixteenth annual meeting of the Berkeley linguistics Society*, 187–204. Berkley: Berkley Linguistics Society.
- Koch, Nikolas & Katharina Günther. 2021. Transfer phenomena in bilingual language acquisition: The case of caused-motion constructions. *Languages* 6(1). 25. <https://doi.org/10.3390/languages6010025>.
- 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*, 441–473. Frankfurt am Main: Peter Lang.
- Kortmann, Bernd, Kerstin Lunkenheimer & Katharina Ehret (eds.). 2020. *The electronic world atlas of varieties of English*, Zenodo. <http://ewave-atlas.org>. (19 May, 2023).
- Krifka, Manfred. 1995. Swahili. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: an international handbook of contemporary research*, vol. 2, 1397–1418. Berlin: De Gruyter.
- Krifka, Manfred. 2003. Semantic and pragmatic conditions for the dative alternation. *Korean Journal of English Language and Linguistics* 4. 1–32.
- Krishnamurti, Bhadriraju. 1985. *A grammar of modern Telugu*. Delhi: Oxford University Press.
- Kroeger, Paul R. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford, CA: CSLI Publications. <https://opac.ku.de/s/uei/de/6/0/BV008272090>.
- Kroll, Judith F. & Paola E. Dussias. 2013. The comprehension of words and sentences in two languages. In Tej K. Bhatia & William C. Ritchie (eds.), *The handbook of bilingualism and multilingualism*, 216–243. Malden, MA: Wiley-Blackwell.
- Kuhn, Max. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* 28(5). 1–26. <https://doi.org/10.18637/jss.v028.i05>.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- 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*

- Massachussets, Amherst on July 24 and 25, 1974*, 77–134. Lisse: Peter de Ridder. <https://doi.org/10.1515/9783110857610-006>.
- Landis, J. Richard & Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1). 159–174. <https://doi.org/10.2307/2529310>.
- Langacker, Ronald W. 1986. An introduction to Cognitive Grammar. *Cognitive Science* 10(1). 1–40. https://doi.org/10.1207/s15516709cog1001_1.
- Langacker, Ronald W. 1999. *Grammar and conceptualization*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110800524>.
- Langacker, Ronald W. 2008. *Cognitive Grammar: A basic introduction*. Oxford: Oxford University Press.
- Laporte, Samantha. 2021. *Corpora, constructions, New Englishes*. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.100>.
- Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry*. The MIT Press 19(3). 335–391.
- Law, Paul. 2016. The syntax of Tagalog relative clauses. *Linguistics* 54(4). 717–768. <https://doi.org/10.1515/ling-2016-0016>.
- Leech, Geoffrey. 1996. *An A-Z of English grammar and usage*. Harlow: Longman.
- Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511642210>.
- Lehmann, Thomas. 1989. *A grammar of modern Tamil*. Pondicherry. <https://opac.ku.de/s/uei/de/6/0/BV004765808>.
- Li, Charles N. & Sandra Thompson. 1989. *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.
- Liao, Yan & Yoshinori J. Fukuya. 2004. Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning* 54(2). 193–226. <https://doi.org/10.1111/j.1467-9922.2004.00254.x>.
- Liberman, Mark. 2014. Samples in which hypercorrections are in. Posting, 15 April, 2014. *Language Log*. <https://languagelog.ldc.upenn.edu/nll/?m=201404&paged=3>. (16 December, 2021).
- Lohse, Barbara, John A. Hawkins & Thomas Wasow. 2004. Domain minimization in English verb-particle constructions. *Language* 80(2). 238–261.
- Lorenz, David. 2020. *To*-contraction in American English: Converging variations and the emergence of horizontal links. In Lotte Sommerer & Elena Smirnova (eds.),

- Nodes and networks in diachronic Construction Grammar*, 243–274.
Amsterdam: John Benjamins. <https://doi.org/10.1075/cal.27.07lor>.
- Lowth, Robert. 1762. *A short introduction to English grammar*. (Ed.) R. C. Alston. In facs. Menston: The Scholar Press.
- MacDonald, Maryellen. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4. 1–16.
<https://doi.org/10.3389/fpsyg.2013.00226>.
- Mair, Christian. 2017. Crisis of the “Outer Circle”? – Globalisation, the weak nation state, and the need for new taxonomies in World Englishes research. In Markku Filppula, Juhani Klemola, Anna Mauranen & Svetlana Vetchinnikova (eds.), *Changing English: global and local perspectives*, 5–24. Berlin: De Gruyter.
<https://doi.org/10.1515/9783110429657-002>.
- Manetta, Emily. 2011. *Peripheries in Kashmiri and Hindi-Urdu: The syntax of discourse-driven movement*. Amsterdam: John Benjamins.
- Maratsos, Michael, Dana E. C. Fox, Judith A. Becker & Mary Anne Chalkley. 1985. Semantic restrictions on children’s passives. *Cognition* 19(2). 167–191.
[https://doi.org/10.1016/0010-0277\(85\)90017-4](https://doi.org/10.1016/0010-0277(85)90017-4).
- Marinis, Theodore, Leah Roberts, Claudia Felser & Harald Clahsen. 2005. Gaps in second language processing. *Studies in Second Language Acquisition* 27(1). 53–78. <https://doi.org/10.1017/S0272263105050035>.
- Marinis, Theodoros & Douglas Saddy. 2013. Parsing the passive: Comparing children with specific language impairment to sequential bilingual children. *Language Acquisition*. Taylor & Francis, Ltd. 20(2). 155–179.
- Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Matthews, Stephen & Virginia Yip. 2011. *Cantonese: A comprehensive grammar*. London: Routledge. <https://opac.ku.de/s/uei/de/6/0/BV039167467>.
- Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, Harald Baayen & Douglas Bates. 2017. Balancing Type I error and power in linear mixed models. *Journal of Memory and Language* 94. 305–315.
<https://doi.org/10.1016/j.jml.2017.01.001>.
- Mazurkewich, Irene. 1985. Syntactic markedness and language acquisition. *Studies in Second Language Acquisition* 7(1). 15–35.
<https://doi.org/10.1017/S0272263100005131>.
- McArthur, Tom. 1987. The English languages? *English Today* 3(3). 9–11.

- McArthur, Tom. 1998. *The English Languages*. Cambridge: Cambridge University Press.
- McDonald, Janet L. 2006. Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language* 55(3). 381–401.
<https://doi.org/10.1016/j.jml.2006.06.006>.
- Mesthrie, Rajend. 2012. Deletions, antideletions and complexity theory, with special reference to Black South African and Singaporean Englishes. In Bernd Kortmann & Benedikt Szmrecsanyi (eds.), *Linguistic complexity: Second language acquisition, indigenization, contact*, 90–100. Berlin: De Gruyter.
- Mesthrie, Rajend & Rakesh M. Bhatt. 2008. *World Englishes: The study of new linguistic varieties*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511791321>.
- 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. <https://doi.org/10.1016/j.jml.2020.104092>.
- Mirdamadi, Farhad Sadri & Nivja H. De Jong. 2015. The effect of syntactic complexity on fluency. *Second Language Research* 31(1). 105–116.
<https://doi.org/10.1177/0267658314554498>.
- Montaut, Annie. 2004. *A grammar of Hindi*. München: Lincom.
- Montgomery, Douglas C. & Elizabeth A. Peck. 1992. *Introduction to linear regression analysis*. New York: Wiley.
- Motus, Cecile L. 1971. *Hiligaynon lessons*. Honolulu: University of Hawaii Press.
- Mukherjee, Joybrato. 2007. Steady states in the evolution of New Englishes: present-day Indian English as an equilibrium. *Journal of English Linguistics* 35(2). 157–187. <https://doi.org/10.1177/0075424207301888>.
- Mukherjee, Joybrato. 2009. The lexicogrammar of present-day Indian English. In Ute Römer & Rainer Schulze (eds.), *Exploring the lexis–grammar interface*, 117–135. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.35.9muk>.
- Mukherjee, Joybrato & Sebastian Hoffmann. 2006. Describing verb-complementational profiles of New Englishes: A pilot study of Indian English. *English World-Wide* 27(2). 147–173. <https://doi.org/10.1075/eww.27.2.03muk>.
- Murakami, Akira, Paul Thompson, Susan Hunston & Dominik Vajn. 2017. ‘What is this corpus about?’: Using topic modelling to explore a specialised corpus. *Corpora*. Edinburgh University Press 12(2). 243–277.

- <https://doi.org/10.3366/cor.2017.0118>.
- Murray, Lindley. 1795. *English grammar: Adapted to the different classes of learners*. (Ed.) R. C. Alston. Repr. In facs. Menston: The Scholar Press.
- Musau, Paul M. 2004. Linguistic human rights in Africa: Challenges and prospects for indigenous languages in Kenya. In Margaret J. Muthwii & Angelina N. Kioko (eds.), *New language bearings in Africa: A fresh quest*, 59–68. Clevedon: Multilingual Matters Ltd.
- Mußemann, Victoria. 2024. *A topic which I want to know more about* – Preposition placement in finite WH-relative clauses in World Englishes. *English Language and Linguistics* FirstView. 1–30.
<https://doi.org/doi:10.1017/S1360674323000667>.
- Nakagawa, Shinichi & Holger Schielzeth. 2013. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4(2). 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>.
- Nelson, Gerald. 1996. The design of the corpus. In Sidney Greenbaum (ed.), *Comparing English Worldwide: The International Corpus of English*, 27–35. Oxford: Clarendon Press.
- Nelson, Gerald. 2002. *The Singapore Corpus: User manual*. Hong Kong.
- Nelson, Gerald. 2017. The ICE project and World Englishes. *World Englishes* 36(3). 367–370. <https://doi.org/10.1111/weng.12276>.
- Nesselhauf, Nadja. 2009. Co-selection phenomena across New Englishes: Parallels (and differences) to foreign learner varieties. *English World-Wide* 30(1). 1–25.
<https://doi.org/10.1075/eww.30.1.02nes>.
- Neukom, Lukas & Manideepa Patnaik. 2003. *A grammar of Oriya*. Zürich: Seminar für Allgemeine Sprachwissenschaft der Universität Zürich.
- Newman, John & Georgie Columbus. 2010. *ICE-Canada: The Canadian component of the International Corpus of English. First release, 2010*. University of Alberta.
- Newman, Paul. 2000. *The Hausa language: An encyclopedic reference grammar*. New Haven: Yale University Press.
- Nykiel, Joanna. 2010. Competence, performance, and extra prepositions. *Journal of English Linguistics*. SAGE Publications Inc 38(2). 143–166.
<https://doi.org/10.1177/0075424209357589>.
- Odlin, Terence. 1989. *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9781139524537>.

- Oehrle, Richard Th. 1976. *The grammar of the English dative alternation*.
Massachusetts Institute of Technology, unpublished PhD diss.
- Okoth-Okombo, Duncan. 1997. *A functional grammar of Dholuo*. Köln: Köppe.
- Omondi, Lucia Ndong'a. 1982. *The major syntactic structures of Dholuo*. Berlin:
Dietrich Reimer.
- Pandharipande, Rajeshwari. 1997. *Marathi*. London: Routledge.
- Paquot, Magali, Jason Grafmiller & Benedikt Szendrői. 2019. Particle placement
alternation in EFL learner vs. L1 speech: Assessing the similarity of
probabilistic grammars. In Andrea Abel, Aivars Glaznieks, Verena Lyding &
Lionel Nicolas (eds.), *Widening the scope of learner corpus research. Selected
papers from the Fourth Learner Corpus Research Conference*, 71–92. Louvai-
la-Neuve: Presses universitaires de Louvain.
- 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, 407–438. Berlin: De Gruyter.
<https://doi.org/10.1515/9783110197181>.
- Payne, John & Rodney D. Huddleston. 2002. Nouns and noun phrases. In Rodney D.
Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge grammar of the
English language*, 323–524. Cambridge: Cambridge University Press.
- Perek, Florent. 2015. *Argument structure in Usage-Based Construction Grammar*.
Amsterdam: John Benjamins.
- Perfetti, Charles A. & Lesley Hart. 2001. The lexical basis of comprehension skill. In
*On the consequences of meaning selection: Perspectives on resolving lexical
ambiguity*, 67–86. Washington D.C.: American Psychological Association.
<https://doi.org/10.1037/10459-004>.
- Pinker, Steven. 1989. *Learnability and cognition: The acquisition of argument
structure*. Cambridge, MA: The MIT Press.
- Pinker, Steven, David S. Lebeaux & Loren Ann Frost. 1987. Productivity and
constraints in the acquisition of the passive. *Cognition* 26(3). 195–267.
[https://doi.org/10.1016/s0010-0277\(87\)80001-x](https://doi.org/10.1016/s0010-0277(87)80001-x).
- Platt, John Talbot, Heidi Weber & Mian Lian Ho. 1984. *The New Englishes*. London:
Routledge & Paul.
- Pollard, Carl & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago:
University of Chicago Press.
- Prat-Sala, Mercè & Holly P. Branigan. 2000. Discourse constraints on syntactic

- processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language* 42(2). 168–182.
<https://doi.org/10.1006/jmla.1999.2668>.
- Pride, John. 1982. *New Englishes*. Rowley: Newbury House Publishers.
- Pullum, Geoffrey K. & Rodney D. Huddleston. 2002. Prepositions and prepositional phrases. In Rodney D. Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge grammar of the English language*, 597–661. Cambridge: Cambridge University Press.
- Pulvermüller, Friedemann, Bert Cappelle & Yury Shtyrov. 2013. Brain basis of meaning, words, constructions, and grammar. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar*, 379–416. Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780195396683.013.0022>.
- Quirk, Randolph. 1985. The English language in a global context. In Randolph Quirk & Henry G. Widdowson (eds.), *English in the world: Teaching and learning the language and literatures*, 1–6. Cambridge: Cambridge University Press.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- R Core Team. 2022. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org>.
- Radford, Andrew. 2004. *Minimalist syntax: Exploring the structure of English*. Cambridge: Cambridge University Press.
- Radford, Andrew. 2019. *Relative clauses: Structure and variation in everyday English*. Cambridge: Cambridge University Press.
- Ray, Tapas S. 2003. Oriya. In George Cardona & Dhanesh Jain (eds.), *The Indo-Aryan languages*, 444–476. London: Routledge.
- Rezai, Mohammad Javad. 2006. Preposition stranding and pied-piping in second language acquisition. *Essex Graduate Student Papers in Language and Linguistics* 8.
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=62a6005a74ad7bb2c8ab445eadcb5862fc9f467c>. (5 February, 2021).
- Richards, Charles Monroe. 1971. *A case grammar of Pampangan. Doctoral dissertation, University of California*. Ann Arbor: University Microfilms International.
- Riedel, Kristina. 2010. Relative clauses in Haya. *ZAS Papers in Linguistics* 53. 211–

226. <https://doi.org/10.21248/zaspil.53.2010.399>.
- Riley, Kathryn & Frank Parker. 1986. Anomalous prepositions in relative clauses. *American Speech* 61(4). 291–306. <https://doi.org/10.2307/454600>.
- Rohdenburg, Günter. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7(2). 149–182. <https://doi.org/10.1515/cogl.1996.7.2.149>.
- Rosenbach, Anette. 2002. *Genitive variation in English: Conceptual factors in synchronic and diachronic studies*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110899818>.
- Rosenfelder, Ingrid, Susanne Jantos, Nicole Höhn & Christian Mair. 2009. *Manual for the Jamaican component (ICE-JA)*. University of Freiburg.
- Ross, John R. 1986. *Infinite syntax!* Norwood, NJ: Ablex Publishing Corporation.
- Röthlisberger, Melanie, Jason Grafmiller & Benedikt Szendrői. 2017. Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics* 28(4). 673–710. <https://doi.org/10.1515/cog-2016-0051>.
- Sadighi, F., M.R. Parhizgar & M. Saadat. 2004. Preposition pied-piping and preposition stranding constructions in the interlanguage grammar of Iranian EFL Learners. *Asian EFL Online Journal* 6(4). Article 3.
- Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics*. Cambridge University Press 33(2). 431–483.
- Sag, Ivan A. 2010. English filler-gap constructions. *Language* 86(3). 486–545.
- Salehi, Mohammad. 2011. The acquisition of preposition pied piping and preposition stranding by Iranian learners of English. *Iranian EFL Journal* 7(3). 85–99.
- Salleh, Ramli. 1989. *Fronted constituents in Malay: Base structures and move alpha in a configurational non-Indo-European language*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Saraceni, Mario. 2010. *The relocation of English: Shifting paradigms in a global era*. London: Palgrave Macmillan. <https://doi.org/10.1007/978-0-230-29691-6>.
- Schachter, Jacquelyn. 1974. An error in error analysis. *Language Learning* 24. 205–214. <http://dx.doi.org/10.1111/j.1467-1770.1974.tb00502.x>.
- Schachter, Paul & Fe. T. Otanes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.
- Schmid, Hans-Jörg. 2010. Does frequency in text instantiate entrenchment in the cognitive system? In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: corpus-driven approaches*, 101–134. Berlin: De

- Gruyter. <https://doi.org/10.1515/9783110226423.101>.
- Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198814771.001.0001>.
- Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*. Manchester. <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>. (26 April, 2021).
- Schmied, Josef. 2004. East African English (Kenya, Uganda, Tanzania): Morphology and syntax. In Bernd Kortmann & Edgar W. Schneider (eds.), *A handbook of varieties of English: a multimedia reference tool*, vol. 2, 929–947. Berlin: De Gruyter.
- Schneider, Edgar W. 2003. The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79(2). 233–281.
- Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511618901>.
- Schneider, Edgar W. 2020. *English around the world*. 2nd edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511781711>.
- Schneider, Gerold & Gaëtanelle Gilquin. 2018. Detecting innovations in a parsed corpus of learner English. In Sandra C. Deshors, Sandra Götz & Samantha Laporte (eds.), *Rethinking linguistic creativity in non-native Englishes*, 47–74. Amsterdam: John Benjamins. <https://doi.org/10.1075/bct.98.03sch>.
- Schwarz, Sarah. 2019. “This must be looked into”: A corpus study of the prepositional passive. *Journal of English Linguistics* 47(3). 249–273. <https://doi.org/10.1177/0075424219851837>.
- Seedorff, Michael, Jacob Oleson & Bob McMurray. 2019. Maybe maximal: Good enough mixed models optimize power while controlling Type I error. PsyArXiv pre-print. <https://doi.org/10.31234/osf.io/xmhfr>.
- Setter, Jane, Cathy S. P. Wong & Brian H. S. Chan. 2010. *Hong Kong English*. Edinburgh: Edinburgh University Press.
- Sharma, Devyani. 2017. English in India. In Laurel J. Brinton & Alexander Bergs (eds.), *The history of English: Varieties of English*, vol. 5, 311–329. Berlin: De Gruyter.
- Shields-Brodber, Kathryn. 1989. Standard English in Jamaica: A case of competing

- models. *English World-Wide* 10(1). 41–53.
<https://doi.org/10.1075/eww.10.1.03shi>.
- Shih, Stephanie, Jason Grafmiller, Richard Futrell & Joan Bresnan. 2015. Rhythm's role in genitive construction choice in spoken English. In Ralf Vogel & Ruben Vijver (eds.), *Rhythm in cognition and grammar*, 207–234. Berlin: De Gruyter.
<https://doi.org/10.1515/9783110378092.207>.
- Shin, Eun Young, Junghyoe Yoon & Taegoo Chung. 2016. Acquisition of preposition stranding and pied-piping in relative clauses and Wh-questions by Korean EFL learners. *Korean Journal of Linguistics* 41(1). 69–95.
- Shin, Eun Young, Junghyoe Yoon & Taegoo Chung. 2017. Preposition pied-piping, stranding, doubling, & dropping in Korean EFL learners' interlanguage. *English Language and Linguistics* 23(2). 89–113.
<http://dx.doi.org/10.17960/ell.2017.23.2.089>.
- Sim, Julius & Chris C. Wright. 2005. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy* 85(3). 257–268.
- Singh, Kandhai & R. Nakkeerar. 2011. *Census of India: Language atlas of India 2011*. New Delhi: Office of the Registrar General & Census Commissioner, India.
https://www.singstat.gov.sg/publications/reference/cop2020/cop2020-sr1/census20_stat_release1. (10 June, 2023).
- Siyanova, Anna & Norbert Schmitt. 2007. Native and nonnative use of multi-word vs. one-word verbs. *International Review of Applied Linguistics in Language Teaching* 45(2). 119–139. <https://doi.org/10.1515/IRAL.2007.005>.
- Sjöholm, Kaj. 1976. A comparison of the test results in grammar and vocabulary between Finnish- and Swedish-speaking applicants for English. In Hakan Ringbom & Rolf Palmberg (eds.), *Errors made by Finns and Swedish-speaking Finns in the learning of English*, ERIC Report ED 122628. Abo: Department of English, Abo Akademi.
- Smirnova, Elena & Lotte Sommerer. 2020. The nature of the node and the network – open questions in diachronic Construction Grammar: Introduction. In Lotte Sommerer & Elena Smirnova (eds.), *Nodes and networks in diachronic Construction Grammar*, 1–42. Amsterdam: John Benjamins.
<https://doi.org/10.1075/cal.27.int>.
- Sommerer, Lotte & Andreas Baumann. 2021. Of absent mothers, strong sisters and peculiar daughters: The constructional network of English NPN constructions. 32(1). 97–131. <https://doi.org/10.1515/cog-2020-0013>.

- Sreekumar, P. 2020. Malayalam. In Sanford B. Steever (ed.), *The Dravidian languages*, 131–161. London: Routledge.
- Sridhar, Shikaripur N. 1990. *Kannada*. London: Routledge.
- Stefanowitsch, Anatol. 2003. Constructional semantics as a limit to grammatical alternation: The two genitives of English. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*, 413–444. Berlin: De Gruyter. <https://doi.org/10.1515/9783110900019.413>.
- Stefanowitsch, Anatol & Susanne Flach. 2017. The corpus-based perspective on entrenchment. In Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*, 101–128. Berlin: De Gruyter. <https://doi.org/10.1515/9783110341423-006>.
- Steger, Maria & Edgar W. Schneider. 2012. Complexity as a function of iconicity: The case of complement clause constructions in New Englishes. In Bernd Kortmann & Benedikt Szmrecsanyi (eds.), *Linguistic complexity*, 156–191. De Gruyter. <https://doi.org/10.1515/9783110229226.156>.
- 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*, 427–448. New York: Psychology Press.
- Street, James A. & Eva Dąbrowska. 2014. Lexically specific knowledge and individual differences in adult native speakers' processing of the English passive. *Applied Psycholinguistics* 35(1). 97–118. <https://doi.org/10.1017/S0142716412000367>.
- Suárez-Gómez, Cristina. 2014. Relative clauses in Southeast Asian Englishes. *Journal of English Linguistics* 42(3). 245–268. <https://doi.org/10.1177/0075424214540528>.
- Suárez-Gómez, Cristina. 2015. Adverbial relative clauses in World Englishes. *World Englishes* 34(4). 620–635. <https://doi.org/10.1111/weng.12165>.
- 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, 1032–1039. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Szmrecsanyi, Benedikt, Jason Grafmiller, Joan Bresnan, Anette Rosenbach, Sali Tagliamonte & Simon Todd. 2017. Spoken syntax in a comparative perspective: The dative and genitive alternation in varieties of English. *Glossa: a journal of general linguistics* 2(1). 1–27. <https://doi.org/10.5334/gjgl.310>.
- 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). 109–137.
<https://doi.org/10.1075/eww.37.2.01szm>.
- Tagliamonte, Sali A. & Jennifer Smith. 2005. No momentary fancy! The zero ‘complementizer’ in English dialects. *English Language & Linguistics*. Cambridge University Press 9(2). 289–309.
<https://doi.org/10.1017/S1360674305001644>.
- Tanzania Government. 2015. *Tanzania Government Portal: Language*.
<https://www.tanzania.go.tz/home/pages/223>. (16 June, 2021).
- Thompson, Katrina Daly & Antonia Fólárín Schleicher. 2001. *Swahili learners’ reference grammar*. Madison: NALRC Press.
- Tomasello, Michael. 1992. *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press.
- Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Traugot, Elizabeth. 2008. The grammaticalization of NP of NP patterns. In Alexander Bergs & Gabriele Diewald (eds.), *Constructions and language change*, 23–46. Berlin: De Gruyter. <https://doi.org/10.1515/9783110211757.23>.
- Trotta, Joe. 2000. *Wh-clauses in English: Aspects of theory and description*. Amsterdam: Rodopi.
- Uhrig, Peter. 2015. Why the Principle of No Synonymy is overrated. *Zeitschrift für Anglistik und Amerikanistik* 63(3). 323–337. <https://doi.org/10.1515/zaa-2015-0030>.
- Uhrig, Peter. 2022. *Large-scale multimodal corpus linguistics – The big data turn*. Friedrich-Alexander-Universität Erlangen-Nürnberg, Habilitation thesis.
- Van de Velde, Freek. 2014. Extending the scope of Construction Grammar. In Ronny Boogaart, Timothy Coleman & Gijsbert Rutten (eds.), 141–180. Berlin: De Gruyter. <https://doi.org/10.1515/9783110366273.141>.
- Van den Eynden, Nadine. 1996. Aspects of prepositional placement in English. In Juhani Klemola, Merja Kytö & Matti Rissanen (eds.), *Speech past and present: Studies in English dialectology in memory of Ossi Ihalainen*, 426–446. Frankfurt am Main: Lang.
- Van Goethem, Kristel & Isa Hendriks. 2021. A diasystematic-constructionist approach: Intensifying constructions in second language acquisition. In Hans C. Boas & Steffen Höder (eds.), *Constructions in contact 2: Language change, multilingual*

- practices, and additional language acquisition*, 375–428. Amsterdam: John Benjamins. <https://doi.org/10.1075/cal.30.12van>.
- Van Rooy, Bertus. 2010. Social and linguistic perspectives on variability in World Englishes. *World Englishes* 29(1). 3–20. <https://doi.org/10.1111/j.1467-971X.2009.01621.x>.
- Van Rooy, Bertus. 2011. A principled distinction between error and conventionalized innovation in African Englishes. In Joybrato Mukherjee & Marianne Hundt (eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*, 189–208. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.44.10roo>.
- Van Rooy, Bertus, Lize Terblanche, Christoph Haase & Joseph Schmied. 2010. Register differentiation in East African English: A multidimensional study. *English World-Wide* 31(3). 311–349. <https://doi.org/10.1075/eww.31.3.04van>.
- Villiers, Jill G. de & Peter A. de Villiers. 1973. Development of the use of word order in comprehension. *Journal of Psycholinguistic Research* 2(4). 331–341. <https://doi.org/10.1007/BF01067055>.
- Vine, Bernadette. 1999. *Guide to the New Zealand component of the International Corpus of English*. Victoria University of Wellington.
- Vitale, Anthony J. 1981. *Swahili syntax*. Dordrecht, NL: Foris Publications.
- Von Eye, Alexander. 2002. *Configural frequency analysis: Methods, models, and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Wali, Kashi. 2005. *Marathi*. München: Lincom Europa.
- Wali, Kashi & Omkar N. Koul. 1997. *Kashmiri: A cognitive descriptive grammar*. London and New York: Routledge. <https://opac.ku.de/s/uei/de/6/0/BV011444265>.
- Wanner, Anja. 2009. *Deconstructing the English passive*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110199215>.
- Ward, Gregory, Betty Birner & Rodney D. Huddleston. 2002. Information packaging. In Rodney D. Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge grammar of the English language*, 1363–1448. Cambridge: Cambridge University Press.
- Wardlow Lane, Liane & Victor S. Ferreira. 2010. Abstract syntax in sentence production: Evidence from stem-exchange errors. *Journal of Memory and Language* 62(2). 151–165. <https://doi.org/10.1016/j.jml.2009.11.005>.
- Wasserscheidt, Philipp. 2014. Constructions do not cross languages: On cross-linguistic

- generalizations of constructions. *Constructions and Frames*. John Benjamins 6(2). 305–337. <https://doi.org/10.1075/cf.6.2.07was>.
- Webb, Victor N. & Kembo-Sure. 2008. The languages of Africa. In Victor N. Webb & Kembo-Sure (eds.), *African voices: An introduction to the languages and linguistics of Africa*, 26–54. Oxford: Oxford University Press.
- Weinreich, Uriel. 1954. Is a structural dialectology possible? *Word* 10(2–3). 388–400. <https://doi.org/10.1080/00437956.1954.11659535>.
- Weinreich, Uriel. 1968. *Languages and contact: Findings and problems*. The Hague: Mouton.
- Wells, Justine B., Morten H. Christiansen, David S. Race, Daniel J. Acheson & Maryellen C. MacDonald. 2009. Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology* 58(2). 250–271. <https://doi.org/10.1016/j.cogpsych.2008.08.002>.
- Whiteley, W. H. & M. G. Muli. 1962. *Practical introduction to Kamba*. London: Oxford University Press. <https://opac.ku.de/s/uei/de/6/0/BV005931114>.
- Williams, Jessica. 1987. Non-native varieties of English: A special case of language acquisition. *English World-Wide*. John Benjamins 8(2). 161–199. <https://doi.org/10.1075/eww.8.2.02wil>.
- Williams, John N. 2006. Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition* 9(1). 71–88. <https://doi.org/10.1017/S1366728905002385>.
- Williams, John N., Peter Möbius & Choonkyong Kim. 2001. Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics* 22(4). 509–540. <https://doi.org/10.1017/S0142716401004027>.
- Winter, Bodo. 2020. *Statistics for linguists: An introduction using R*. New York: Routledge.
- Wong, Jock. 2014. *The culture of Singapore English*. Cambridge: Cambridge University Press.
- Wulff, Stefanie. 2016. A friendly conspiracy of input, L1, and processing demands: *That*-variation in the language of German and Spanish learners of English. In Andrea Tyler, Lourdes Ortega, Hae In Park & Mariko Uno (eds.), *The usage-based study of language learning and multilingualism*. Georgetown: Georgetown University Press.
- Wulff, Stefanie & Stefan Th. Gries. 2019. Particle placement in learner language.

- Language Learning* 69(4). 873–910. <https://doi.org/10.1111/lang.12354>.
- 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. <https://doi.org/10.1515/9783110572186-004>.
- Wulff, Stefanie, Nicholas Lester & Maria T. Martinez-Garcia. 2014. *That*-variation in German and Spanish L2 English. *Language and Cognition* 6(2). 271–299. <https://doi.org/10.1017/langcog.2014.5>.
- Xiao, Richard. 2009. Multidimensional analysis and the study of World Englishes. *World Englishes* 28(4). 421–450. <https://doi.org/10.1111/j.1467-971X.2009.01606.x>.
- York, Richard. 2012. Residualization is not the answer: Rethinking how to address multicollinearity. *Social Science Research* 41(6). 1379–1386. <https://doi.org/10.1016/j.ssresearch.2012.05.014>.
- Zehentner, Eva & Elizabeth Closs Traugott. 2020. Constructional networks and the development of benefactive ditransitives in English. In Lotte Sommerer & Elena Smirnova (eds.), *Nodes and networks in diachronic Construction Grammar*, 167–211. Amsterdam: John Benjamins. <https://doi.org/10.1075/cal.27.05zeh>.
- Zeldes, Amir. 2013. Productive argument selection: Is lexical semantics enough? *Corpus Linguistics and Linguistic Theory* 9(2). 263–291. <https://doi.org/10.1515/cllt-2013-0006>.
- Zhang, Shi. 1990. Correlations between the double object construction and preposition stranding. *Linguistic Inquiry*. The MIT Press 21(2). 312–316.
- Zorc, R. David. 1977. *The Bisayan dialects of the Philippines: Subgrouping and reconstruction*. Canberra: Australian National University. <https://doi.org/10.15144/PL-C44>.
- Zuur, Alain F., Elena N. Ieno & Chris S. Elphick. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1(1). 3–14. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>.
- 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*, 101–142. New York: Springer.