STUDIES IN ENGLISH LANGUAGE

Thomas Hoffmann

Preposition Placement in English

A Usage-based Approach

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521760478

This page intentionally left blank

Preposition Placement in English

Preposition placement, the competition between preposition-stranding (*What is he talking about?*) and pied-piping (*About what is he talking?*), is one of the most interesting areas of syntactic variation in English. This is the first book to investigate preposition placement across all types of clauses that license it, such as questions, exclamations and *wh*-clauses, and those which exhibit categorical stranding, such as non-*wh* relative clauses, comparatives and passives. Drawing on over 100 authentic examples from both first-language (British) and second-language (Kenyan) data, it combines experimental and corpus-based approaches to provide a full grammatical account of preposition placement in both varieties of English. Although written within the usage-based Construction Grammar framework, the results are presented in theory-neutral terminology, making them accessible to researchers from all syntactic schools. This pioneering volume will be of interest not only to syntacticians, but also to second-language researchers and those working on variation in English.

THOMAS HOFFMANN is Assistant Professor of English Linguistics at the University of Osnabrück, Germany. His main research interests are Construction Grammar and the syntactic and phonetic variation in World Englishes. He has co-edited the volume *World Englishes: Problems, Properties and Prospects* (2009; with Lucia Siebers) and together with Graeme Trousdale is currently co-editing the *Oxford Handbook of Construction Grammar.*

STUDIES IN ENGLISH LANGUAGE

The aim of this series is to provide a framework for original studies of English, both present-day and past. All books are based securely on empirical research, and represent theoretical and descriptive contributions to our knowledge of national and international varieties of English, both written and spoken. The series covers a broad range of topics and approaches, including syntax, phonology, grammar, vocabulary, discourse, pragmatics and sociolinguistics, and is aimed at an international readership.

Already published in this series:

Charles F. Meyer: *Apposition in Contemporary English*
Jan Firbas: *Functional Sentence Perspective in Written and Spoken Communication*
Izchak M. Schlesinger: *Cognitive Space and Linguistic Case*
Katie Wales: *Personal Pronouns in Present-Day English*
Laura Wright: *The Development of Standard English, 1300–1800: Theories, Descriptions, Conflicts*
Charles F. Meyer: *English Corpus Linguistics: Theory and Practice*
Stephen J. Nagle and Sara L. Sanders (eds.): *English in the Southern United States*
Anne Curzan: *Gender Shifts in the History of English*
Kingsley Bolton: *Chinese Englishes*
Irma Taavitsainen and Päivi Pahta (eds.): *Medical and Scientific Writing in Late Medieval English*
Elizabeth Gordon, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury and Peter Trudgill: *New Zealand English: Its Origins and Evolution*
Raymond Hickey (ed.): *Legacies of Colonial English*
Merja Kytö, Mats Rydén and Erik Smitterberg (eds.): *Nineteenth-Century English: Stability and Change*
John Algeo: *British or American English? A Handbook of Word and Grammar Patterns*
Christian Mair: *Twentieth-Century English: History, Variation and Standardization*

Evelien Keizer: *The English Noun Phrase: The Nature of Linguistic Categorization*
Raymond Hickey: *Irish English: History and Present-Day Forms*
Günter Rohdenburg and Julia Schlüter (eds.): *One Language, Two Grammars?: Differences between British and American English*
Laurel J. Brinton: *The Comment Clause in English*
Lieselotte Anderwald: *The Morphology of English Dialects: Verb Formation in Non-standard English*
Geoffrey Leech, Marianne Hundt, Christian Mair and Nicholas Smith: *Change in Contemporary English: A Grammatical Study*
Jonathan Culpeper and Merja Kytö: *Early Modern English Dialogues: Spoken Interaction as Writing*
Daniel Schreier, Peter Trudgill, Edgar Schneider and Jeffrey Williams: *The Lesser-Known Varieties of English: An Introduction*
Hilde Hasselgård: *Adjunct Adverbials in English*
Raymond Hickey: *Eighteenth-Century English: Ideology and Change*
Charles Boberg: *The English Language in Canada: Status, History and Comparative Analysis*

Preposition Placement in English

A Usage-based Approach

THOMAS HOFFMANN

University of Osnabrück

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521760478

First published 2011

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Hoffmann, Thomas, 1976–
Preposition Placement in English : A Usage-based Approach / Thomas Hoffmann.
p. cm. – (Studies in English Language)
Includes bibliographical references and index.
ISBN 978-0-521-76047-8
1. English language–Prepositions. 2. English language–Syntax. 3. English language–Grammar. I. Title. II. Series.
PE1335.H63 2011
428.2–dc22
2010045708

ISBN 978-0-521-76047-8 Hardback

Contents

Figures

Tables

Abbreviations

AdjP	Adjective Phrase
ANOVA	Analysis of Variance
AVM	Attribute-Value Matrix
BIRM	Birmingham corpus
BNC	British National Corpus
BROWN	Brown Corpus
CFA	Configural Frequency Analysis
CP	Complementizer Phrase
DP	Determiner Phrase
FTF	Fuzzy Tree Fragment
GHFP	Generalized Head Feature Principle
GLM	Generalized Linear Model
HCFA	Hierarchical Configural Frequency Analysis
HPSG	Head-driven Phrase Structure Grammar
ICE	International Corpus of English
ICECUP	ICE-GB retrieval software
ICE-EA	International Corpus of English (East Africa)
ICE-GB	International Corpus of English (Great Britain)
ICN	Instant Chunk Number
LF	Logical Form
LLC	London–Lund Corpus
MCN	Mean Chunk Number
ME	Middle English
NP	Noun Phrase
OE	Old English
P	Preposition

PASS	Passive
PP	Preposition Phrase
PRO	Pronoun
PST	Past (tense)
REL	Relative
SpecC	Specifier of Complementizer Phrase
SpecFocus	Specifier of Focus Phrase
SpecT	Specifier of Tense Phrase
SpecTopic	Specifier of Topic Phrase
V	Verb
VP	Verb Phrase
XP	Unspecified phrase (NP, VP, AdjP, etc.)

Acknowledgements

> L – – d! said my mother, what is all this story about?——
>
> A COCK and a BULL, said Yorick —— And one of the best of its kind, I ever heard. (Laurence Sterne, *The Life and Opinions of Tristram Shandy, Gentleman*)

Hopefully the present book does not turn out to be a 'cock and bull' story but qualifies as valid scientific analysis of preposition placement in English. Also I would not go as far as labelling it 'one of the best of its kind', but the current version is definitely better than any of the earlier drafts. This has to do with the great number of friends and colleagues who constantly provided invaluable feedback and support.

First and foremost, I have to thank Edgar W. Schneider for his continuing support and friendship throughout the various stages of this project. It was he who suggested preposition-stranding as a potential topic for an MA thesis seven years ago, and little did we both realize at the time that it would eventually lead to a PhD thesis and, finally, end up as a book in CUP's *Studies in English Language* series. The relationship between my MA and PhD theses can probably be best described as loose family resemblance: both deal with preposition-stranding, but while the former limited itself to relative clauses, the PhD investigated all constructions exhibiting the phenomenon in present-day English. Furthermore, in contrast to the MA study, for the PhD thesis I corroborated corpus data with psycholinguistic experiments and focused not only on British English but compared it to the second-language variety of English spoken in Kenya. Finally, while the syntactic description of preposition placement in my MA thesis was couched in arcane Minimalist Program terminology, the PhD opted for an arcane Construction Grammar analysis. From this short overview of the numerous changes that the work has undergone in the past few years it will be obvious that supervising my theses was not an easy task. Edgar, however, put up with all these linguistic mood swings. In addition to having him as my thesis supervisor, for the past five years I have also had the pleasure of working for Edgar's research team. Both as a supervisor and as a boss he not only allowed me a great deal of freedom to pursue my own linguistic research, but also offered support whenever I needed it (be it for grant applications or comments on earlier drafts of articles). For all of this, I can't thank him enough and I am very much aware how lucky I am to have him as a boss and, even more importantly, as a friend.

Next I would like to thank my second supervisor, Roswitha Fischer. Way back in the last millennium when I was still a student she was the first to offer me a job as a student tutor at the English department of the University of Regensburg. At the time I mainly wanted the job for financial reasons, but it was clearly the first step towards my decision to pursue an academic career in English linguistics. I don't know whether this has done the field of linguistics much good, but I certainly profited from this decision. So I sincerely would like to thank Roswitha for making all this possible in the first place.

As usual, there are many others without whom nothing even closely resembling the present book could have been written. During my field trip to Kenya, Alfred Buregeya from the University of Nairobi helped me find safe accommodation, ensured that I got enough subjects for all of my experiments and invited me to his house to spend some time with his wonderful family. Graeme Trousdale and And Rosta, on the other hand, supplied me with a large number of my British English subjects for the experiments. I am deeply grateful to all three of them for their help. In addition to that, I would like to thank Stefan Th. Gries for providing me with his Coll.analysis 3 software, and I am also very grateful to John Maindonald for his help with the R software, as well as to John Paolillo and Robert Sigley for their Goldvarb support. I am furthermore indebted to Sam Featherston for introducing me to the WebExp software and for his continuous help with all questions concerning the statistical analysis of Magnitude Estimation data in SPSS. Also, I am in Manfred Sailer's debt for having a look at my earlier endeavours into HPSG. Andreas Dufter I would like to thank for proofreading chapter 3. A big thank you also goes to Holger Saurenbach, who not only provided me with constant food for thought, but in essence converted me to Construction Grammar. Finally, I have to thank my friend Peter Waller who agreed to proofread my manuscript. I hope I can return the favour in the not so distant future.

Next, I would like to thank everyone at Cambridge University Press for their permanent support in publishing this book. In particular, I am, of course, extremely grateful to the series editor, Merja Kytö, for considering my book a possible contribution to the excellent *Studies in English Language* series in the first place. On top of this, I would like to express my gratitude to the commissioning editor, Helen Barton, for her continuous support throughout the entire publication process. Finally, I would like to thank the two anonymous reviewers for their valuable comments, which have helped to greatly improve the quality of the book.

Last but not least, I would like to thank my wife Moni, my rock of love, who's keeping it real and always knows what time it is.

1 Introduction

1.1 Preposition placement: The need for corroborating evidence

> A corpus and an introspection-based approach to linguistics are not mutually exclusive. In a very real sense they can be gainfully viewed as being complementary. (McEnery and Wilson 1996: 16)

It is well known that linguistic generalizations based on corpus data face two potential problems: (1) just because a phenomenon cannot be found in a corpus, it cannot be concluded that it is ungrammatical (the 'negative data' problem), and (2) just because a construction appears in a corpus it does not automatically follow that it is grammatical (the 'performance' problem). Introspective grammaticality judgements, on the other hand, are not flawed by these problems but the sentence stimuli used in such studies (1) have to be invented by the researcher (the 'unnatural data' problem) and (2) thus do not allow the investigation of contextual factors such as the level of formality (the 'context' problem). As the quote above shows, this complementary nature of corpus and grammaticality judgement data leads McEnery and Wilson to argue for a combination of both methods, instead of choosing one over the other.

While many researchers still tend to draw on either corpus or introspection data, I have argued repeatedly (cf. Hoffmann 2006, 2007a) that the approach suggested by McEnery and Wilson can yield insights well beyond what the two data sources would allow individually. In this book I will show how treating carefully collected and interpreted introspection and corpus data as 'corroborating evidence' (cf. Hoffmann2006; an approach that independently has been called 'converging evidence' by Gries, Hampe and Schönefeld 2005) can be used to shed light on a particularly complex area of syntactic variation within the English language, namely preposition placement.

In English relative clauses, for example, a preposition can either precede the *wh*-relativizer ('preposition pied-piping',[1] see (1.1a)) or it can appear

[1] The term was coined by Ross in analogy to the children of Hamlin who followed the pied piper in the well-known fairy tale (Ross 1986: 126, n. 23).

without an adjacent NP complement in front of the relativized position '$\_{i}$' ('preposition-stranding' (cf. 1.1b)).

(1.1) a. I want a data source [on which]$_i$ I can rely $\_{i}$
b. I want a data source [which]$_i$ I can rely on $\_{i}$

Linguistic phenomena such as (1.1) which exhibit two or more variants can be investigated as 'dependent variables' in quantitative language variation studies. The basic underlying principle of such approaches is that the choice of a particular variant ('pied-piped' or 'stranded') of a dependent variable ('preposition placement') will be influenced by factors such as its linguistic context, stylistic level and social identity (so-called 'independent variables'; cf. Preston 1996: 2; Sigley 1997: 19). With respect to the distribution of the dependent variable 'preposition placement' in relative clauses there is already considerable disagreement within the literature: opinions range from 'stranding is not really an option with WH-... relatives' (Van den Eynden 1996: 444) to the claim that stranding is used fairly frequently 'in spoken English, standard as well as non-standard' (Bergh and Seppänen 2000: 295). Yet, *wh*-relatives are only one of the many contexts in which a preposition can be stranded or pied-piped.

Wh-relatives are classic examples of so-called 'filler–gap' constructions (cf. Pollard and Sag 1994: 157), i.e. sentences in which a constituent in a non-argument position (the 'filler', i.e. [*on which*]$_i$ in (1.1a) and [*which*]$_i$ in (1.1b)) has been displaced from the position in which it would normally appear in a declarative sentence (cf. *I rely on this data source*). This normal position, with which the filler is still logically associated, is called a 'gap' (indicated by '$\_{i}$' in (1.1)). Other such filler–gap constructions are, for example, *wh*-questions or topicalized clauses, and these are also contexts which license variable preposition placement (cf. (1.2) and (1.3), respectively):

(1.2) a. [On what]$_i$ can I rely $\_{i}$?
b. [What]$_i$ can I rely on $\_{i}$?
(1.3) a. [On this data source]$_i$ you can rely $\_{i}$.
b. [This data source]$_i$ you can rely on $\_{i}$.

In addition to this, there are also other clausal contexts in which an element in argument position is associated with a stranded preposition, such as passives (1.4) or 'hollow clauses' (i.e. 'non-finite clauses ... other than relatives or open interrogatives where some non-subject NP is missing but recoverable from an antecedent NP or nominal', Huddleston 2002c: 1245; cf. (1.5)):

(1.4) a. [Pied-piping]$_i$ has been talked about$_i$ enough.
b. *[About pied-piping]$_i$ has been talked$_i$ enough.
(1.5) a. [His thesis]$_i$ was easy [to find fault with$_i$].
b. *[With his thesis]$_i$ was easy [to find fault$_i$].

As the above examples show, in cases where the associated element functions as an argument (as subject in (1.4) and (1.5)), no pied-piped alternative is possible.

The above examples illustrate that preposition placement is clearly affected by clause type. Besides this, various other independent factors such as the syntactic function of the prepositional phrase (PP), the type of phrase in which a PP is contained (whether the PP is embedded in a verb (VP), adjective (AdjP) or noun phrase (NP)), the level of formality or even processing factors have been claimed to restrict the stranding/pied-piping alternation (cf. e.g. Bergh and Seppänen 2000; Gries 2002; Hoffmann 2005; Trotta 2000). Yet, while all earlier accounts of preposition placement only focused on specific clause types, the present book attempts to investigate the distribution of preposition pied-piping and stranding in all of the possible clause types.

1.2 World Englishes, usage-based linguistics and preposition placement

What is often referred to as 'the' English language is in fact a heterogeneous and linguistically fascinating group of first (L1), second (L2), pidgin and creole as well as foreign language varieties (cf. e.g. Kortmann *et al.* 2004; Mesthrie and Bhatt 2008). In light of this, it is somewhat surprising that virtually all previous empirical studies on preposition placement only restricted themselves to a description of the phenomenon in Standard British or American English, in particular since such an approach makes it difficult to disentangle variety-specific phenomena from general linguistic constraints.

Now one way to overcome this problem would obviously be to carry out a large-scale comparative study of preposition placement across many different World English varieties. However, such an approach introduces a great number of new variables, such as possible L1 influence on L2 Englishes as well as the effect of formal English language teaching in the respective countries (which is of particular relevance for preposition placement since in Britain or the US e.g. the pied-piped variant used to be endorsed at school as the correct choice in formal text types). Considering that on top of this, no full-scale empirical analysis exists that takes into account all variables (including e.g. all different clause types) affecting preposition placement in any of the classic standard varieties of English, a different approach was chosen for the present book.

Instead of a large-scale comparison, it was decided to focus on an in-depth analysis of preposition-stranding and pied-piping in L1 British English and L2 Kenyan English, with the latter variety being chosen for the following reasons:

- First of all, English in Kenya is a stable L2 variety: it is commonly used as a lingua franca by speakers of different native languages and is also employed in parliament and government institutions. Moreover, English is 'the language of secondary and tertiary education and the High Court' (Hudson-Ettle and Schmied 1999: 4; cf. also Kanyoro 1991: 406; Schmied 2004a: 924; Skandera 2003: 20).

- Next, it was possible to gather data that was fully comparable to the British English one: as part of the International Corpus of English (ICE) project (Greenbaum 1996) comparable corpora have been compiled for both British and Kenyan English, from which all relevant stranded and pied-piped tokens could be extracted. On top of this, it was possible to collect introspection data for both varieties using Magnitude Estimation experiments (cf. Bard, Robertson and Sorace 1996).
- Moreover, the most widely used Kenyan textbooks were identified (*Head Start Secondary English*: Bukenya *et al.* 2003a, 2003b; Bukenya, Kioko, and Njeng´ere 2004, 2005; and *New Integrated English*: Gathumbi *et al.* 2003, 2004a, 2004b, 2005) and examined with respect to the advocated position on preposition placement.
- Finally, drawing on the existing literature it was possible to fully assess the L1 influence of the local lingua franca Swahili (Barrett-Keach 1985; Brauner and Bantu 1967; Vitale 1981) as well as the other local languages, i.e. Bantu languages such as Kikuyu (Leaky 1959), Kamba (Whitely and Muli 1962) and Luyia (Appleby 1961; Donohew 1962) or Nilotic languages like Dholuo (Omondi 1982).

Thus, while any two-way comparison of varieties might have its limitations, this approach allows an in-depth analysis of all intra- as well as extra-linguistic variables that are identified as potentially affecting preposition placement.

However, as I will show, the empirical investigation of preposition placement in a classic first-language variety such as British English and a second-language variety such as Kenyan English does not only advance our knowledge of the grammatical differences between these two varieties of English. Such an approach also allows identification of general cognitive principles affecting preposition-stranding and pied-piping and to explore their interaction with input frequency effects. For while general processing principles should affect first- and second-language speakers alike, input frequency in L2s might be limited due to a restriction of the variety in question to certain functional domains. In Kenya, for example, English is used in many official domains (such as education or government), but in informal, personal situations speakers are often more likely to draw on one of their local L1s (cf. Schmied 2004a: 923–4). As so-called 'usage-based' approaches (cf. Barlow and Kemmer 2000; Bybee 2006; Bybee and Hopper 2001; Langacker 1987, 2005; Tomasello 2003) have pointed out, however, input plays a crucial role in shaping our mental grammars. In fact, recent research has shown that all linguistic levels from phonology (cf. e.g. Bybee 2000, 2001; Pierrehumbert

2001) to morphology (cf. e.g. Bybee 1985, 1995; Hay and Baayen 2005) and syntax (cf. e.g. Casenhiser and Goldberg 2005; Saffran, 2001, 2002; Stefanowitsch and Gries 2005) are heavily affected by input frequency effects: every time a word is encountered, it leads to the activation of patterns of neural nodes in the mind. The strength of the connections of these neural nodes is thus directly affected by the word's input frequency (also known as 'token frequency'). The more often a word is used, the stronger the association of the neural nodes will become, essentially leading to long-term mental storage. Once an item is stored in this way, it is said to be cognitively entrenched (see Croft and Cruse 2004: 292–3; Langacker 1987: 59–60). Yet, input frequency does not only affect the storage of words, it also plays a role in the entrenchment of abstract grammatical patterns: structures with a high type frequency, i.e. those that have been encountered with many different lexicalizations (such as *John gave Bill a book*, *Peter sent Mary a letter*, *She forwarded him the mail*), all of which share a common meaning ('A causes B to receive C by V-ing'), can lead to the entrenchment of abstract grammatical patterns (such as Subject_A V Object_B Object_C; Goldberg 2006: 39; cf. also Bybee 1985, 1995; Croft and Cruse 2004: 308–13; Goldberg 2006: 98–101).

Now, a syntactic theory which explicitly allows the incorporation of such usage-based input effects as well as general processing factors is Construction Grammar (see e.g. Croft 2001; Fillmore and Kay 1996; Ginzburg and Sag 2000; Goldberg 2003), which was one of the reasons why this framework was adopted for the present theory. Recently, various different Construction Grammar approaches have been proposed (e.g. Croft's (2001) Radical Construction Grammar or Goldberg's (2006) Cognitive Construction Grammar), all of which share the fundamental idea that all grammatical, including syntactic, knowledge is stored mentally as constructions (i.e.form–meaning pairings). Thus Construction Grammarians assume that abstract clausal patterns such as Subject_A V Object_B Object_C 'A causes B to receive C', are stored form–meaning pairings, i.e. constructions, just like simple words such as *apple* or *man*. The only difference is that the latter have a fixed phonological form, and are therefore called 'substantive' constructions, while the former, an example of a 'schematic' construction, consists of slots that can be filled by various lexical items (such as *John gave Bill a book* or *Peter sent Mary a letter*; see e.g. Croft and Cruse 2004: 247–9). These two types of constructions then represent the end points of a lexicon–syntax cline from fully substantive to fully schematic constructions (cf. Croft and Cruse 2004: 255; Goldberg 2003: 220; Jackendoff 2002: 176). Examples of partly-filled, partly-open constructions falling in between these endpoints would be e.g. idioms such as [Subject_A *kick*-TENSE[2] *the bucket*] 'die(A)' (cf.

[2] 'TENSE' is shorthand notation for a link to the various independent tense constructions such as [HAVE V-*en*] 'Present Perfect' or [*will* V] 'Future', which will specify the final inflected form of the verb as well as the presence of any auxiliaries.

John kicked the bucket / *His pal has kicked the bucket* or *She will kick the bucket*) or the comparative correlative construction [*the* X-*er*, *the* Y-*er*] 'if X then Y' (e.g. *the more I drink, the better I sing* or *The less you say, the better*; Jackendoff 2002: 172–87).

The particular version of Construction Grammar approach which I shall advocate in this book is a usage-based version of Ginzburg and Sag's (2000) Head-Driven Phrase Structure Grammar (HPSG) approach. HPSG is normally not a usage-based framework, but it employs a full-fledged grammar formalism that is both explicit and falsifiable and has been widely used for the description of a great number of grammatical phenomena. On top of that, its constraints can easily be reinterpreted as constructions (cf. Ginzburg and Sag 2000; Sag 1997).[3] As I will show, the HPSG approach thus allows us to model computationally how individual constructions combine, something that at present is not possible to the same degree of explicitness in other Construction Grammar accounts (e.g. Croft 2001; Fillmore and Kay 1996; Goldberg 2006). Yet, in light of the results on the influence of input on mental grammars mentioned above as well as corroborating evidence for this from language acquisition (Diessel 2006; Diessel and Tomasello 2000; Lieven *et al.* 2003), I deem it absolutely crucial that usage-based phenomena are also taken into account. I will therefore also illustrate how usage-based information can be incorporated into HPSG-Construction Grammar approaches. In particular I will investigate the possibility that first-language speakers possess more substantive as well as abstract schematic constructions in their mental grammar than second-language learners, since, as pointed out above, the latter normally receive much less input of the target language than native speakers do. Furthermore, following Hawkins (2004), I take it that for both L1 and L2 speakers processing factors play an important role in the formation of abstract schemata. If the same content can be expressed by two competing structures and one of these is easier to process than the other (a claim that has been made for pied-piping, for example; see Deane 1992; Gries 2002; Hawkins 1999, 2004), then the simpler structure will be preferred in performance. Consequently, it will be used more often with a greater range of lexicalizations, which increases its type frequency and ultimately leads to it being more cognitively entrenched than its alternative

[3] Something that has led Sag himself to develop an HPSG-based Construction Grammar approach called Sign-Based Construction Grammar (SBCG; Sag 2007). SBCG, however, like HPSG is not usage-based. Instead, SBCG and HPSG are representative of so-called 'complete inheritance' Construction Grammar models (cf. Croft and Cruise 2004: 276–8). Complete inheritance approaches aim to limit the number of constructions postulated for a language to an absolute minimum that still allows a speaker to generate combinatorially all grammatical structures. On top of that, such approaches usually also employ constructions that are just abstract schemas without a paired meaning (cf. e.g. the Subject–Auxiliary inversion SAI construction; Fillmore 1999). Other proponents of complete inheritance Construction Grammar models include Jackendoff (2002; cf. also Culicover and Jackendoff 2005) or Fillmore and Kay (1996).

(cf. Hawkins 2004: 6). Finally, competition between two structures such as pied-piping and stranding also can lead to 'preemption' (Goldberg 2006: 99) playing an important role: if on a particular occasion one construction is used instead of a potential alternative, then the hearer will assume that this choice reflects a functional difference between the two structures. Ultimately, this will lead to the functional differentiation of the two alternatives (though as I will try to show this seems to be an effect that is stronger for L1 speakers, since these receive more input which allows constructions to extend beyond their prototypical context expected by processing constraints).

Due to its combination of empirical data analysis of preposition placement in L1 British and L2 Kenyan English with a usage-based syntactic framework, the present study should be of interest to linguists working on syntactic variation and varieties of English as well as second-language researchers. Besides this, the main readership will probably consist of syntacticians, especially those working within a Construction Grammar framework. In order to make the empirical results accessible to researchers from other frameworks, however, the Construction Grammar analysis of the data will be deferred to the last chapter and the preceding discussion of all empirical data will be presented in a terminology as theory-neutral as possible. How these chapters tie in with the overall structure of the book will be discussed next.

1.3 Outline

The book is divided into seven chapters. After this introductory chapter, chapter 2 ('Corroborating evidence: Data and methodology') argues that linguists should not restrict themselves to either corpus or introspection data. Elaborating on the idea of corroborating evidence (cf. Hoffmann 2007a), I claim that both types of data can be collected and interpreted in an objective, reliable and valid way. The chapter then presents the corpora employed in the present study (the British English and Kenyan English components of the International Corpus of English (ICE) project) as well as the statistical tools used for the analysis of tokens displaying categorical (*Coll.analysis 3* and *HCFA 3.2* scripts for R for Windows: Gries 2004a, 2004b) or variable effects (*Goldvarb*: Robinson, Lawrence and Tagliamonte 2001; *Rbrul*: D. Johnson 2009b). Finally, the experimental method (Magnitude Estimation) for the elicitation of introspection data and the details of their statistical analysis are introduced (cf. Bard et al. 1996; Cowart 1997).

Following this, chapter 3 ('Case notes: Independent factors') gives an overview of the various factors that have been claimed to influence preposition placement in English. These include clause type (3.1), type of PP (3.2), level of formality (3.3), NP- vs VP-/AdjP-embedded PPs (3.4), processing complexity (3.5) and, finally, second-language-specific ones (3.6).

Chapter 4 ('Evidence I: Corpus results') then presents the results from the statistical analysis of the two ICE corpora. On the one hand, these show that preposition placement in Kenyan English is subject to the same kinds of processing constraints as in British English (with e.g. prepositional verbs favouring stranding far more than locational adjuncts). On the other hand, variety-specific effects can also be identified (in that, for example, less formal contexts exhibit a strong preference for stranding in relative clauses in British English, while in Kenyan English pied-piping is preferred in relative clauses regardless of the level of formality).

In chapter 5 ('Evidence II: Experimental results') the results from three Magnitude Estimation experiments are discussed (two on relative clauses and one on interrogative clauses). Again, processing factors (the type of PP) show similar effects in both varieties, as do grammatical constraints (in relative clauses pied-piping with *that*, e.g. **the man on that I relied*, is significantly judged worse than with *who*, e.g. **the man on who I relied*). In addition to this, variety-specific effects can be observed as well (e.g. pied-piping with prepositional verbs being less entrenched in Kenyan English).

Chapter 6 ('Preposition placement: The case for a Construction Grammar account') then provides a Construction Grammar account of the empirical findings, arguing that two general constructions (i.e. a 'stranded' and a 'pied-piped' one) are not enough to account for the great number of categorical and variable factors affecting preposition placement in the two varieties. Instead, a usage-based HPSG Construction Grammar analysis is put forward that is computationally unproblematic and, on top of that, allows incorporating usage-based information (in that statistically significant collocations such as *way in which* in *the way in which I did it* can be said to be stored in the speaker's mental grammar). Finally, chapter 7 ('Conclusion: The verdict') sums up the results of the study.

As mentioned above, the theoretical analysis of preposition placement has deliberately been deferred to the end of the book. Researchers working within different syntactic frameworks can thus access the empirical findings of the present study without constantly having to worry about unfamiliar theoretical concepts or notations. For while I take Construction Grammar approaches to be both observationally maximally adequate as well as psychologically plausible, I consider it of prime importance to provide an empirically adequate description of the data in question that can be evaluated by a maximum number of my peers. This, however, seems to imply that the linguistic community has already agreed on which data to employ and how to interpret findings based on these data. As the next chapter will show, however, this is not at all the case.

2 Corroborating evidence: Data and methodology

2.1 'What counts as evidence in Linguistics'[1]

As Penke and Rosenbach point out, 'nowadays most linguists will probably agree that linguistics is indeed an empirical science' (2004: 480). However, while the importance of empirical evidence is generally acknowledged by most researchers, the following quotations from Sampson and Chomsky show that there is no agreement among linguists as to the type of data that is to be analysed empirically:

We do not need to use intuition in justifying our grammars, and as scientists, we must not use intuition in this way. (Sampson 2001: 135)

You don't take a corpus, you ask questions … You can take as many texts as you like, you can take tape recordings, but you'll never get the answer. (Chomsky in Aarts 2000: 5–6)

If both Sampson's position on introspection and Chomsky's views on corpora were correct, there would obviously be no valid data base left for linguists to investigate. Fortunately, however, Sampson and Chomsky are only extreme proponents of their respective schools of linguistics. Nevertheless, when investigating a particular syntactic phenomenon, many linguists still only draw on either corpus or introspection data (though there seems to be an increasing number of exceptions such as Gries, Hampe and Schönefeld 2005, the collected volume by Kepser and Reis 2005 or the special issue on corpus and experimental techniques of *Corpus Linguistics and Linguistic Theory* 5.1 – in particular see Gilquin and Gries 2009). In the literature this preference for either of the two types of data is often attributed to different epistemological approaches (e.g. Lemnitzer and Zinsmeister 2006: 14–32).

Linguists like Sampson are said to be influenced by empiricism, a philosophical school which advocates the prime importance of experience and favours an inductive scientific approach. Followers of Chomsky, on the other hand, are said to be influenced by rationalism, which emphasizes rational hypothesizing and is characterized by a deductive approach. While the

[1] Penke and Rosenbach (2004: 480).

preference for a particular type of data by an individual school might be explained by its philosophical background, I consider this fact immaterial for the present study. Instead, I claim that in order to qualify as scientific evidence it must only be ensured that a specific linguistic data type meets the major constraints normally imposed on empirical research, i.e. that data

(1) must be objective, i.e. interpersonally observable (cf. Sampson 2001: 124),
(2) allow for valid and reliable measurements (cf. Kline 1993).

As I will show, while the criticism of a specific type of data is not always couched in these terms, it is in fact the objectivity, validity and reliability of introspection and corpus data that is questioned by proponents of the alternative schools of linguistics.

Interestingly, advocates of both introspection and corpus data usually follow the same line of argument: the weaknesses of introspection/corpus data are x, y and z. Yet x, y and z are exactly the advantages of the competing methodology. That is why you should never use the former, but only stick to the latter type of data.

The argument for introspection data usually runs like this: corpora only exhibit a speaker's 'performance', which is influenced by 'memory limitations, distractions, shifts of attention and interest and errors (random and characteristic)' (Chomsky 1965: 3). Thus a speaker's performance, i.e. corpus data, is only an indirect and partly flawed reflection of his competence. As a result, corpus data are haunted by the 'performance' problem: just because a sentence appears in a corpus doesn't mean that it is grammatical. In addition to this, it is generally accepted that linguistic competence enables a speaker to create an infinite number of sentences. Yet, how should a finite corpus contain all the examples relevant for the analysis of a particular problem (cf. McEnery and Wilson 1996: 4–10)? This obviously leads to the well-known 'negative data' problem: just because a construction does not surface in a corpus it does not follow that it is ungrammatical. Therefore, the intuition of a native speaker drawing on his competence has to be preferred over the examination of corpus data.

The argument for corpus data, on the other hand, usually runs like this: the sentences used for introspective judgements are 'unnatural', invented data which lack a communicative context. Judgements on these sentences are then collected in an unsystematic, unscientific way: most of the time the linguist will only rely on his or her own intuitions. Thus linguists who use introspective data 'produce theory and data at the same time' (Labov 1972: 199). If anyone then casts doubts on their judgements, these linguists resort to the claim that judgements might vary but that in their idiolect the sentence is in fact grammatical/ungrammatical (Sampson 2001: 137). Since introspection thus yields data which cannot be refuted, it must be considered 'un-scientific' (e.g. Sampson 2001: 124). Finally, the intra- and inter-speaker stability of introspection data is questioned (for an overview, see Cowart 1997: 4f.) and sometimes speakers even say things that they believe sincerely that they would never produce (Sampson 2001: 136). Due to these problems one should always stick to authentic data provided by corpora.

A closer look at the above arguments reveals that they imply that corpus and introspection data violate the objectivity, validity and reliability constraints on empirical research.

The main criticism of corpus data obviously concerns their validity. An empirical measurement technique can only be ‘said to be valid if it measures what it claims to measure’ (Kline 1993: 15). Now the opponents of corpus data claim that the object of study in linguistics is the linguistic competence of a speaker. Since corpus data are ‘flawed’ by performance factors, they do not constitute a valid means of investigating linguistic competence. Before evaluating this argument, it needs to be pointed out that not only corpus data are measurements of linguistic performance. As Schütze (1996: 6, 14–15) has argued, introspective judgements are also subject to performance factors. Thus, introspection offers ‘a different access path from language use to competence, [but is itself] just another sort of performance’ (Schütze 1996: 6). Accordingly, if performance was in fact only a flawed mirror of competence, neither corpus nor introspective data could be considered valid.

Yet, as Leech noted, ‘the putative gulf between competence and performance has been overemphasised’ (1992: 108), since the latter clearly is the product of the former. Following Occam’s razor, it would therefore be much more reasonable to assume that, under normal circumstances, performance should actually be a rather good mirror of a speaker’s competence. Second, the guiding principle of modern corpus design is representativeness: nowadays, corpora are designed as statistically representative samples of populations (cf. McEnery and Wilson 1996: 63–6). Although a corpus can never contain all sentences of a language, it will at least be a carefully constructed miniature model. Thus, both corpus data and introspection can be considered valid ‘access paths’ to competence. Nevertheless, the two are obviously not measuring exactly the same phenomenon: while the strength of the former lies in constituting positive data which can be analysed for frequency and context effects, the latter allows the investigation of the grammaticality of negative data. It is this complementary nature of the validity of these different types of data that calls for them to be treated as corroborating evidence. Note that complementing introspection data with corpus data also has the advantage of minimizing the one problem affecting the validity of the former: since ‘informant judgements don’t always agree with the informant’s own linguistic behaviour’ (Cowart 1997:5; cf. e.g. Labov 1975), corpus data can be used to check the validity of subject’s grammaticality judgements.

As can be seen above, one of the main criticisms of introspection data concerns their reliability, i.e. their intra- and inter-subject consistency (cf. Kline

1993: 5). While informal methods of introspection data elicitation might indeed fail to be reliable, recently several studies have provided statistical proof that judgements elicited via carefully constructed experiments are in fact intra- and inter-subject consistent (cf. Bard, Robertson and Sorace 1996; Cowart 1997: 12–28; Keller 2000). The reliability of corpus data – at least with respect to test–retest reliability (Kline 1993: 5) – on the other hand, has never been questioned: since most corpora, with the exception of monitor corpora, are finite samples, repeated measurements of a phenomenon automatically yield reliable results. Yet, the question of intra- and inter-subject consistency also applies to corpus data: however balanced and representative a corpus might be, there is no guarantee that all informants have contributed the same number of tokens of a particular linguistic phenomenon. In fact, apart from highly frequent phenomena (such as NPs) it is very likely that some informants might not have produced any relevant data while others have contributed two or more tokens to a corpus. As long as it can be entertained that linguistic theory should be 'concerned primarily with an ideal speaker-listener, in a completely homogenous speech community' (Chomsky 1965: 3f.), this obviously does not constitute a problem: if linguistic knowledge were in fact distributed homogeneously across a speech community, then it would not matter how many tokens an individual contributed to a corpus. However, while the mutual intelligibility of speakers of a single dialect does indeed imply a certain degree of homogeneity within a speech community, ever since Labov's groundbreaking New York study (1966) it has become apparent that all speech communities display variation on all linguistic levels (for an overview see e.g. Hudson 1996; Milroy and Gordon 2003; Trudgill 1995). Consequently, since not all informants contribute an equal amount of data on a particular phenomenon, it cannot automatically be assumed that overall corpus results reflect the linguistic behaviour of all informants – let alone of all speakers of the speech community in question.

This problem is even more acute for second-language-learner corpora: it is well known that language learners differ greatly with respect to the rate at which and the degree to which they attain a second language (cf. e.g. R. Ellis 1995: 99–126, 1998: 73–78; Granger 1998: 5–6; Leech 1998: xviii). Therefore if a more advanced learner adds little or no data to a corpus while a less proficient one contributes a great number of tokens, the resulting skewed data set clearly can not be considered representative of all second language learners. Now one way to preclude such a confounding effect is of course the choice of a representative set of informants which one expects to be linguistically as homogeneous as possible. In addition to this, however, it is also necessary to test statistically whether the data set in question is in fact homogeneous. This requires statistically analysing corpus data, ensuring that the resulting statistical models have a good fit for the data and minimizing the undue influence of individual informants (in the case of preposition-stranding and pied-piping by e.g. subjecting corpus data to a Generalized Linear Model

(GLM) analysis, checking the model fit and the error structure of the final GLM model and establishing the cross-validation accuracy of the model; see Maindonald and Braun 2003: 209f. and section 2.2 below for details). Finally, another way to ensure the intra- and inter-subject consistency of results based on corpus data is to corroborate them in an introspection experiment in which all subjects provide equal information on all tested conditions. In such experiments the homogeneity of the subjects' judgements – and via statistical inference the homogeneity of the speech community – can thus explicitly be tested.

The fact that both types of data can in fact be tested in a reliable way is also important for the third criterion of empirical research mentioned above: objectivity. Since results of empirical studies should be independent of a particular researcher, it is obviously essential that repeated measurements of a data source by different researchers yield similar results. Objectivity in turn guarantees that results obey the meta-principle of empirical research: the fact that conclusions drawn from data analysis must be refutable/falsifiable. As Sampson correctly emphasizes: 'All that matters is that any feature of the theory which is doubted can be confirmed or refuted on empirical grounds' (2001: 137). As mentioned above, this is indeed a problem if in the face of competing judgements, a linguist argues that in his/her idiolect a construction is (un-)grammatical. While this may or may not be the case, it is a virtually unfalsifiable claim. Again it must be stressed that nowadays an increasing number of linguists refrain from only relying on their own introspection. Instead introspection data is collected obeying the standards of psychological experiments: via the selection of a representative sample of subjects, careful design of experimental materials, randomization of the order of stimuli, the use of fillers/distractors and the employment of statistical analysis (cf. Bard, Robertson and Sorace 1996; Cowart 1997; Schütze 1996). Clearly, introspection data collected in such a way can be considered both objective and falsifiable. Since corpora easily allow hypotheses to be tested by different researchers, the objectivity and falsifiability of this data source is also uncontroversial.

Thus, both corpus and introspection qualify as good data sources for empirical research. Both are valid, reliable and objective data which allow for the generation of falsifiable hypotheses. Their individual strengths derive from the fact that they are measuring slightly different phenomena: corpora yield good data for the statistical analysis of positive data with respect to frequency and context phenomena. Single tokens or the absence of a construction in a corpus are, of course, interesting and important findings. Any hypothesis based on such data, however, needs to be corroborated by experimental introspection data in order to ensure empirical validity. Introspection, on the other hand, allows the investigation of negative data and, given the right experimental method, an elicitation of subtle grammaticality judgements. In addition to this, it should be noted that due to the creative aspect of language – a classic argument for this type of data – introspection experiments can also profit from corpus studies when it comes to test design: if the set of possible sentences is infinite, how should a linguist be expected to come up with all relevant examples of a construction? Or as Jan Aarts puts it: ‘Only linguists who use corpus data themselves will know that a corpus always yields a much greater variety of constructions than one can either find in the literature or think up oneself’ (1991: 46). Thus, a corpus study can also yield data for further experimental studies which otherwise would not have been detected. (For while linguistic competence might be homogeneous in a speech community – a claim which requires much further empirical research – creativity clearly varies greatly from one individual to another.) As a result, I disagree with Sampson and Chomsky, and instead completely agree with McEnery and Wilson, who ask

> [w]hy move from one extreme of only natural data to another of only artificial data? Both have known weaknesses. Why not use a combination of both, and rely on the strengths of each to the exclusion of their weaknesses? A corpus and an introspection-based approach to linguistics are not mutually exclusive. In a very real sense they can be gainfully viewed as being complementary. (1996: 16)

Elsewhere (Hoffmann 2006), I have called the use of multiple sources of different types of data ‘corroborating evidence’ (an approach independently advocated as ‘converging evidence’ by Gries, Hampe and Schönefeld 2005). The basic idea behind this term is that, just like in a criminal investigation, linguists have to amass enough corroborating evidence to convince a jury of their peers. Accordingly, for the present study I will draw on introspection and corpus data to come up with a case on preposition placement in British and Kenyan English that will hopefully convince my peers.

In a criminal case, however, it is also important that the adduced evidence is gathered and interpreted in a forensically sound way. For the present study I take this to mean that, whenever possible, statistically analysed quantitative data are preferred over ‘hearsay’, i.e. qualitative data. The latter might no doubt provide interesting insights, but I argue that the former helps to present a much stronger case. In the following chapters I will elaborate on my various sources of evidence as well as the forensic tools used, i.e. the statistical analyses with which the data were interpreted.

2.2 Exhibit A: Corpus data

When contrastively investigating two or more varieties it is crucial to ensure that the data which are to be analysed have been sampled in a comparable way. Only then is it possible to claim that differences between samples actually reflect differences between the varieties in question (e.g. Leech 2007: 141–2).

For corpus-based contrastive studies on varieties of English it is therefore of vital importance to employ 'matching' or 'comparable corpora', i.e. 'a set of two or more corpora whose design differs, as far as possible, in terms of only one parameter; the temporal or regional provenance of the textual universe from which the corpus is sampled' (Leech 2007: 141–2).

For the synchronous varieties of English around the world the International Corpus of English (ICE) project is such an attempt to compile a set of comparable corpora. The project was initiated by Sidney Greenbaum in 1988, and in addition to British English, ICE corpora have, for example, been compiled for varieties such as East African (Kenyan and Tanzanian English), Hong Kong, Indian, Irish, Jamaican, New Zealand, Philippines or Singaporean English. Based on identical design principles, all these corpora employ a common annotation scheme 'in order to ensure maximum comparability' (Nelson, Wallis and Aarts 2002: 3). Furthermore, all corpora will not only be tagged for part-of-speech but will also be parsed for syntactic structure. At present, however, only the British English component, the ICE-GB corpus, is available fully tagged and parsed.

Since all other ICE corpora, including the East African one (ICE-EA), are to be modelled on the British English component, I will now first give an overview of the ICE-GB before detailing the specific properties of the ICE-EA corpus.

2.2.1 *The International Corpus of English (ICE) corpora*

2.2.1.1 *ICE-GB*

ICE-GB was 'compiled and grammatically analysed at the Survey of English Usage, University College of London, between 1990 and 1998' (Nelson, Wallis and Aarts 2002: 3), with all texts dating from 1990 to 1993 inclusive. It is a one-million-word corpus, consisting of spoken (about 637,000 words) as well as written (about 423,000 words) material. Every text in the corpus has been assigned a unique text code which identifies its text category (e.g. S1A-001 to S1A-090 are face-to-face conversations; see Nelson, Wallis and Aarts 2002: 309–31). Concerning their stylistic level, both spoken and written texts in ICE-GB range from less formal (private face-to-face conversations and social letters) to rather formal (public legal cross-examinations and printed academic writings). Furthermore, some categories, such as dialogues and student examination scripts, are obviously produced more spontaneously than printed texts, which allow more planning time and have undergone an extensive editorial process (Nelson, Wallis and Aarts. 2002: 5ff.).

All the speakers and writers in the corpus are adults (age 18 or over), and, with few exceptions, were born in England, Scotland or Wales. All informants have completed secondary-level schooling, with many having 'received tertiary education as well' (Nelson, Wallis and Aarts 2002: 5). ICE-GB is thus intended to be a representative sample of educated British English in the 1990s (Nelson, Wallis and Aarts 2002: 4f.). As a result, although the corpus contains texts from all levels of formality (from private conversations to academic writings), all data from the ICE-GB corpus undeniably constitute samples of the most educated end of the British English sociolect. If, as is sometimes claimed, the pied-piping preference is indeed to a great extent dependent on education (McDaniel, McKee and Bernstein 1998: 309), then, regardless of the level of formality, the current study can be said to investigate the sociolect with the greatest pied-piping tendency.

ICE-GB is fully tagged for part-of-speech and parsed for syntactic structure.[2] Thus each word in the corpus has been assigned a word-class label ('tag') and in many cases also additional features, providing information such as (e.g.) a verb's transitivity. Furthermore, its syntactic function has also been identified. (For an overview of the various tags, features and syntactic functions encoded in ICE-GB see Nelson, Wallis and Aarts. 2002: 22–68.)

(2.1) I very much enjoyed the work that I was involved in<PS, PREP>

The word *in* in (2.1),[3] for example, carries the tag 'PREP', which means that it has been classified as a preposition (Nelson, Wallis and Aarts 2002: 34–5). Furthermore, if a preposition is not followed by a complement it is assigned the syntactic function 'PS (stranded preposition)' (cf. *in* in (2.1); see also Nelson, Wallis and Aarts 2002: 53). With the help of the ICE-GB's retrieval software (a program called ICECUP) it was thus possible to extract all relevant stranded tokens for the present study via the stranded preposition 'PS'-tag. Note however that prepositions which are part of incomplete utterances (2.2) or unintelligible fragments (2.3) also do not exhibit an overt complement, and accordingly these are also parsed as 'stranded prepositions' in the corpus. The output of the ICECUP query thus still had to be inspected manually and tokens such as (2.2) and (2.3) were excluded from further analysis.

(2.2) It 's **[like]** <ICE-GB:S1A-053 #257:1:C>

(2.3) <unclear> three-or-four-words </unclear>[4] soft **[in]** <unclear> <unclear-words> </unclear> <ICE-GB:S1A-018 #57:1:B>

[2] For detailed information about the tagging and parsing procedure, see Nelson, Wallis and Aarts (2002: 13–17).

[3] For expository purposes the tags of the other words in (2.1) have been omitted.

[4] All information from the ICE-GB corpus contained within angle brackets < > represents part of the corpus's structural markup, i.e. additional information added by the compilers. In (2.3), for example, the string '<unclear> three-or-four-words </unclear>' means that there were three or four words in the text which could not be identified by the compilers. A complete list of all structural markup symbols employed in ICE-GB can be found in Nelson, Wallis and Aarts (2002: 333).

The second advantage of ICECUP which facilitated the extraction of the relevant data from ICE-GB was that ICECUP has a so-called 'Fuzzy Tree Fragment' (FTF) option which allows the user to search the corpus for abstract syntactic structures (cf. Nelson, Wallis and Aarts 2002: 116ff.). Instead of having to limit the search for pied-piped constructions to, for example, specific preposition+*wh*-word constructions (e.g. 'in which', 'to which', 'of which', etc.), it was therefore possible to design FTFs which for a given *wh*-item found all instances in the corpus where it was governed by a preposition (i.e. 'P + which', 'P + who', etc.).

In light of the above features of ICE-GB, it should have become apparent that the corpus constituted the perfect data source for the present study: it includes texts from all levels of formality and is the model based on which several comparable corpora have been or are currently being compiled. In addition to this, the accompanying software ICECUP allowed the retrieval of all relevant stranded and pied-piped data from the corpus. Unfortunately, as I will show next, the source of the Kenyan corpus data, the East African component of the ICE project (ICE-EA), does not possess several of these advantages.

2.2.1.2 ICE-EA

The Kenyan corpus data were retrieved from the ICE-EA corpus, which consists of spoken and written texts from Kenya and Tanzania, all dating from between 1990 and 1996 (cf. Hudson-Ettle and Schmied 1999: 5). In line with the general ICE sampling scheme, all informants are 'adults (over 18) who have received formal education up to at least secondary level schooling through the medium of English or have a public status that makes their inclusion appropriate' (Hudson-Ettle and Schmied 1999: 5). Thus concerning the criteria of the sampled time period as well as the included informants the ICE-EA corpus can be said to match ICE-GB.

In total, the Kenyan subcorpus of the ICE-EA corpus comprises 791,695 words (289,625 words in the spoken component + 100,207 words in the written- as-spoken section and401,863 words from written texts; see Hudson-Ettle and Schmied 1999: 53–63). The ICE-EA sampling scheme differs somewhat from that of ICE-GB. The reason for this is that the specific situation in East Africa required several changes in order to guarantee the representativeness of the data (as being typical of Kenyan and Tanzanian English; see Schmied 1996). In particular, it turned out to be impossible to sample the text types described below, all of which feature in the ICE-GB corpus (cf. Hudson-Ettle and Schmied 1999: 4).

While the spoken part of ICE-GB contains 'phonecalls', 'business transactions' and 'unscripted monologues', none of these could be sampled for either Kenyan or Tanzanian English. While it was also impossible to record unscripted 'legal presentations', the researchers at least were able to obtain handwritten versions of such texts, which they included in the written component. Finally, no recordings from parliament or the courts could be made in Kenya or Tanzania (i.e. there are no 'parliamentary debates' or 'legal cross-examinations' texts in the ICE-EA corpus), but transcripts of spoken dialogue from both institutions in Kenya were made available to the researchers. Since these transcripts had been composed by third parties and not by the ICE-EA team, it was decided to create a new category in addition to 'spoken' and 'written', the 'written-as-spoken' component (Hudson-Ettle and Schmied 1999: 6–8). Finally, unlike the written section of ICE-GB, 'skills/hobbies' texts do not appear to be of particular cultural importance in East Africa, so that no texts were sampled for this category.

As the above discussion shows, while the ICE-EA is modelled as closely as possible on the ICE-GB corpus to ensure comparability, the situation in Kenya and Tanzania required several changes to the sampling scheme to ensure the representativeness of the data (see Schmied 1996). Nevertheless, due to their balanced and comparable composition the ICE-GB and the ICE-EA corpora can be considered the most reliable data sources available for any contrastive study on British and Kenyan English.

The greatest disadvantage of the ICE-EA corpus for the present study was that so far it has not been tagged or parsed. Yet, in order to allow the corpus results to be generalized to the underlying population it was considered essential to retrieve all instances of stranded and pied-piped prepositions. The only way to guarantee full retrieval meant printing out and manually reading through the entire Kenyan subcorpus.

2.2.2 *Forensic tools I: Goldvarb, R, Coll.analysis 3 and HCFA*

After having described the corpus evidence which has been investigated for the present study, I will now present the forensic tools, i.e. the programs used for its statistical analysis.

The choice of the correct statistical test for a set of data greatly depends on what type of response/dependent variable one is investigating and what kind of explanatory/independent variables one suspects of affecting the response/dependent variable (see Crawley 2005: 1–2). In the main corpus studies I carried out (see chapter 4), the response / dependent variable 'preposition placement' was binary, i.e. it only had two variants, 'stranded' or 'pied-piped'. Furthermore, all explanatory variables were categorical, i.e. had two or more levels but no continuous intermediate levels (for the apparent exception of the factor complexity, see section 3.5). For such cases the correct statistical test is a subtype of the Generalized Linear Model called binary logistic regression analysis (cf. Baayen 2008: 195–208; Crawley 2005: 2, 117–18, 269–80; Gries 2008: 284–94; Paolillo 2002). The statistical research tool deemed appropriate for the present study was the Goldvarb 2001 computer program for Windows[5] (Robinson, Lawrence and Tagliamonte 2001), which provides descriptive statistical information as well as a multivariate binary logistic regression analysis of the data.

Now while Goldvarb might seem like the straightforward choice to readers with a sociolinguistic background, many people familiar with other statistical packages often argue against using the program for a number of reasons (cf. e.g. Szmrecsanyi 2006: 88–9). Before addressing the main points of criticism, let me stress that I see statistical analysis as a tool and not an end in itself. Furthermore, to be honest, my main motivation for using Goldvarb was that preliminary analyses on parts of the British English data had already been carried out with this program (Hoffmann 2005, 2006). It is true that Goldvarb is a single-purpose software that only carries out binary logistic regression, while software suites such as the commercial program SPSS or the free, open-source program R (R Development Core Team 2008) offer a much greater choice of statistical tests and come with much better graphical facilities. In fact, this wider range of options, together with the recent publication of a number of great introductory textbooks on statistics with R for linguists (such as Baayen 2008; Gries 2008; K. Johnson 2008), probably means that in the future I will abandon Goldvarb altogether and opt for R instead. Nevertheless, from a statistical point of view all that matters for the present book is that the corpus data required a binary logistic regression analysis, which is exactly the analysis that Goldvarb provides. (Though, as the following discussion will show, all Varbul models were also checked in R and the information from these analyses is included in the Appendix for readers who are more familiar with logistic regression in R.)

One point of criticism that is occasionally levelled at Goldvarb is the fact that it reports its parameters on a different scale from most other statistical software.[6] Logistic regression parameters can be reported on either the logit or the probability scale (Paolillo 2002: 162). While other statistical programs, such as SPSS or R, report parameters on the logit scale (ranging from $-\infty$ to ∞ with a neutral value of 0), Goldvarb gives factor weights on the probability scale. As a result, the neutral value for Goldvarb factors is 0.5, with factors ranging from 0 to <0.5 having an inhibiting and those from >0.5 to 1 having a favouring influence on the investigated variant of the dependent variable. Yet, since the logistic function is the inverse of the logit function, Goldvarb parameters share the same properties as the SPSS/R logit parameters (Paolillo 2002: 160–2): values that are equidistant from the neutral value in either direction have the same magnitude of effect (thus Goldvarb factors of

[5] Source: www.york.ac.uk/depts/lang/webstuff/goldvarb.

[6] E.g. an anonymous reviewer objected to the use of Goldvarb, stating, amongst other issues, that the program's 'outputs are less natural to interpret than log odds'.

0.2 and 0.8 have the same magnitude of effect, but in the opposite direction). Furthermore, 'small differences in parameter values have a smaller effect around the neutral value' (Paolillo 2002: 162; i.e. around 0.5 in Goldvarb and around 0 in SPSS/R). On top of that, Goldvarb's probability weights can easily be transformed into logit coefficients (or to be more precise sum contrast log odds) via the formula ln((p/1-p)). Thus Goldvarb weights of 0.4 and 0.6 correspond to logit log odds of -0.405 and + 0.405, respectively (D. Johnson 2009a: 361). I understand that, depending on individual preferences, people might find one or the other way of reporting regression coefficients/factor weights easier to interpret.[7] In order to facilitate the interpretation of the logistic regression analyses, it was therefore decided to re-run the Goldvarb analysis in R using Rbrul (www.ling.upenn.edu/~johnson4/Rbrul.R; D. Johnson 2009a, b). Amongst other options, Rbrul emulates Goldvarb analyses in R and outputs the coefficients on the probability as well as the logit scale. Consequently, all logistic regression models presented in this book will give the effects of independent factors as Varbrul probability weights as well as logit log odds. This should enable all readers to interpret the effects of independent factors, regardless of whether they have a Varbrul or SPSS/R background.

Next, independent variables with a categorical effect (e.g. contexts which obligatorily induce stranding or pied-piping) are always problematic for binary logistic regression analyses. Goldvarb, for example, cannot compute such categorical effects or 'knockout constraints' at all (Young and Bayley 1996: 272–4; Sigley 1997: 240). Consequently, tokens exhibiting such factors either have to be eliminated from the data, or grouped together ('recoded') with other non-categorical factors from the same factor group, provided there are sufficient linguistic reasons supporting such a regrouping[8] (Paolillo 2002; Sigley 1997: 240; Young and Bayley 1996: 272–4). Again this might be construed as a specific disadvantage of Goldvarb since SPSS, R and other software packages can calculate coefficients for such categorical effects (cf. Baayen 2008: 196). However, the parameters for knockout effects reported by these programs can be unreliable and are often accompanied by various warning messages (this phenomenon is usually discussed as (quasi-) complete separation in the statistical literature; see e.g. Allison 2008; Field 2009: 274–5; Hosmer and Lemeshow 2000: 138–40). Thus the issue of how to deal with knockout constraints is one that has to be addressed by all researchers using logistic regression, regardless of the software they are working with.

[7] Note, however, that it's not only people working with Varbrul that find probabilities easier to interpret than logits (see e.g. Gelman and Hill 2007: 80).

[8] A third possibility would have been to add a 'fictitious token' (cf. Paolillo 2002) coded only for the categorical environment, and for the dependent variant. However, such a fictitious token might distort model results if the categorical environment is not very frequent (cf. Hoffmann 2005).

Even more importantly, from a linguistic point of view, I agree with Daniel Johnson (2009b) that truly 'invariant contexts … should be excluded from an analysis of variation'. As I will show, preposition placement is constrained by several such categorical effects (e.g. obligatory stranding with *that*-relativizers: cf. *the data that he talked about* vs **the data about that he talked*). In cases where it was suspected that knockout effects were due to genuine categorical constraints, the affected tokens were therefore dropped from the logistic regression analysis. In order to be able still to further investigate these constraints and their interaction with other variables, alleged categorical effects were then tested experimentally and the corpus data in question were subjected to various other statistical tests (see below).

Szmrecsanyi (2006: 88) voices two further points of criticism of Goldvarb, namely that the program 'cannot easily handle interactions' and 'does not report the extent of variance explained'. While I consider the former unproblematic (since all that matters is that interactions can be modelled at all; see e.g. Sigley 2003), the latter is definitely more serious: even if the best logistic regression model has been identified, it still has to be investigated how well this model actually fits the data. Model selection in Goldvarb proceeds via 'step-up/step-down regression analysis', which tests the contribution of the individual factor groups for significance via a G^2-test (the standard test for maximum likelihood models which can also be used to assess whether a recoded model with fewer parameters is as good as the original, more complex model; see Crawley 2005: 103–24; Paolillo 2002: 140–2; Sigley 2003). Once the perfect, 'minimal adequate model' (Crawley 2005: 104) for a data set has been found, Goldvarb offers two parameters that indicate whether it is a good fit for the data: (1) the 'Fit: X-square' test (a G^2-test that checks whether the maximum possible, i.e. the log-likelihood of the data, and the model log-likelihood can be said to approximate each other) as well as (2) a chi-square value for the differences of the actual and expected realizations for each cell created by all factor combinations ('Error') and the overall chi-square for all cells. Szmrecsanyi, however, is right that none of these parameters specifies the amount of variation accounted for by the model. Fortunately, however, Rbrul calculates an R^2 value for logistic regression models (namely Nagelkerke-R^2) which allows one to quantify the amount of variation explained by the best Goldvarb model (with Nagelkerke-R^2 ranging from 0 to 1, which correspond to 0 per cent and 100 per cent of variation explained by a model, respectively).

For all logistic regression models presented in this book I will thus report raw frequencies, probability and log odds logit coefficients as well as model fit parameters (Nagelkerke-R^2 and the 'Fit: X-square' test). On top of this, it will also be checked whether the number of included factors is justified considering the sample size. Following Sigley (2003: 251), the threshold value for the maximum number of *S* parameters per *n* number of tokens used in this study is that of Freeman (1987): $n > 10\,(S+1)$. While this value is likely to be too optimistic for linguistic studies (since it assumes complete independence of tokens), it at least provided a maximum threshold value against which the results of the corpus studies could be compared. As long as the number of employed parameters was considerably lower than the resulting Freeman threshold, a model was deemed fully acceptable.

Finally, as pointed out in section 2.1, apart from the fit of a model it is also important to preclude the undue effect of individual tokens and to guarantee the homogeneity of the data on which a model is based. Unfortunately, this is something that indeed cannot be tested in Goldvarb. For this reason, the best model identified by Goldvarb was always fed into the R 2.7.1 for Windows software (R Development Core Team 2008).[9] This made it possible to compute the so-called cross-validation parameter (cf. Maindonald and Braun 2003: 121–3, 209–10). This test assesses the predictive accuracy of a model by randomly splitting up the data into a number of subsets (so-called 'folds'; I always use Maindonald and Braun's 2003 *cv.binary()* function for this test, which by default creates ten folds). Each fold then becomes a test set against which the model's accuracy is assessed. Only if this procedure yields a high value for the cross-validation parameter can it be guaranteed that a model fits well to all of the created folds. Consequently this procedure helps to ensure that individual tokens as well as undue influence of data from single speakers can be factored out.

As a final note on the logistic regression chapter, let me say that I am aware that some readers might find it excessive that I thus report the findings from both Goldvarb and R analyses. Nevertheless, as pointed out above, this approach should make the results accessible to researchers from both traditions (and, since the statistical output of both R and Goldvarb are provided in the Appendix, allow them to check the validity of the proposed models).

Next, while the above statistical tools were used to test the overall distribution of preposition placement, it was also deemed important to check whether specific syntagmatic lexicalizations were significantly associated with particular syntactic constructions. Take for example the relative clause structure in (2.4):

(2.4) I like the way in which he did it.

As will be seen in section 4.4.2, *way in* is actually the most frequent antecedent–preposition syntagm in manner adjunct relative clauses in the ICE-GB. This raises the question, however, of whether the two words are also significantly associated – or in other words whether there is statistical support for the claim that *way in*-manner relative clauses are a lexically stored syntagm for British English speakers. In order to answer questions such as these I will draw on Stefanowitsch and Gries's covarying-collexeme analysis

[9] For more information go to: cran.r-project.org.

Table 2.1 *Covarying-collexeme analysis (adapted from Stefanowitsch and Gries 2005: 9)*

	in in slot2 (word *in* in slot 2)	¬ *in* in slot2 (all other words in slot 2)
way in slot1 (word *way* in slot 1)	Frequency (*way* + *in*)	Frequency (*way* + ¬*in*)
¬*way* in slot1 (all other words in slot 1)	Frequency (¬*way* + *in*)	Frequency (¬*way* + ¬*in*)

(2005: 9–11). Table 2.1 shows that the underlying data structure of such a covarying-collexeme analysis is always a two-by-two table. The association which is to be tested is the grey shaded cell in Table 2.1, i.e. the syntagm *way in*. Now in order to statistically assess the association of *way* and *in* in manner relative clauses this syntagm must be compared to all other logically possible lexicalizations of manner relative clauses: i.e. the frequency of the antecedent *way* occurring with another preposition than *in* (*way* + ¬*in*), the frequency of *in* co-occurring with another antecedent noun (¬*way* + *in*), as well as the frequency of all other antecedent noun and preposition combinations (¬*way* + ¬*in*).

Normally, the standard test applied by linguists to two-by-two tables such as Table 2.1 is the chi-square test. However, chi-square tests always require the expected frequencies of all cells to be greater than 5 (Woods, Fletcher and Hughes 1986: 144f.), a requirement that is often violated by collocational data such as Table 2.1. In the covarying-collexeme analysis the association of two slots of a construction is therefore instead tested via the Fisher-Yates Exact test. Unlike chi-square tests, the Fisher-Yates Exact test does not require large frequencies and places no particular restrictions on its input data (see Baayen 2008: 113). The only disadvantage of the Fisher-Yates Exact is that it is computationally expensive and thus extremely time-consuming when calculated manually. Yet this could be avoided by using the *Coll.analysis 3* script (Gries 2004a), which allows for the automatic calculation of the Fisher-Yates Exact test via R.

For the interpretation of the output of a covarying-collexeme analysis it is important to note that the *Coll.analysis 3* script does not report simple *p*-values for the association of two words, but log-transformed *p*-values. Thus 'values with absolute values exceeding 1.30103 are significant at the level of 5% (since $10^{-1.30103} = 0.05$), and values exceeding 2 and 3 are significant at the levels of 1% and 0.1% respectively'[10] (Stefanowitsch and Gries 2005: 7). One reason for the log-transformation of the *p*-values is that 'the most interesting values are only located in the small range of 0.05 and 0 (and

[10] Since log(0.01) = −2 and log(0.001) = −3.

many linguists are unfamiliar with the scientific format employed for representing such small numbers)' (Stefanowitsch and Gries 2005: 7). Another is that unlike *p*-values, the base-ten logarithm of the *p*-value (called 'p_{log10}') can be more easily interpreted as a measure of association strength, especially since in the covarying-collexeme analysis the sign of the resulting p_{log10} value is changed 'to a plus when the observed frequency is higher than the expected one' (Stefanowitsch and Gries 2005: 7). Thus a p_{log10} value of −2 indicates that a pair of words appears significantly less often in a construction than expected, while a p_{log10} value of 2 implies that the two words in question are significantly associated (both results being significant at a *p*-value of 0.01).

Finally, as was pointed out above, Goldvarb cannot compute categorical factors. Yet, it would, of course, be interesting to see whether specific categorical contexts (say, prepositional passives, which only allow stranding; cf. *He was talked about* vs **About he was talked*) exhibit different preferences in British and Kenyan English with respect to the PP types they license (e.g. whether affected-location PPs such as *this bed has been slept in* are significantly more frequent in British than in Kenyan English). In cases such as these when the correlation of three or more nominal variables was investigated, the data in question were subjected to a 'configural frequency analysis' (CFA; Bortz, Lienert and Boehnke 1990: 155–7; Gries 2008: 242–54).

In a CFA each combination of factors is labelled a 'configuration'. Consequently, a factor arrangement like British English × affected location PP × Passives would qualify as a configuration. Table 2.2 gives a (fictitious) example of the output of a CFA. It shows that for each configuration of factors Factor _1 × […] × Factor _n (BritishE × […] × Passives and KenyanE × […] × Passives in the table) the observed frequency ('Frequ') as well as its expected frequency ('Exp') is given. Based on these a specific chi-square ('Cont.chisq') value is calculated. So far, a CFA thus resembles a normal chi-square test. Yet, while in a simple chi-square test only a single test is performed on the data, in a CFA the significance of each configuration is calculated ('P.adj.bin'). Since multiple tests are thus carried out over the same data set, the significance values for each configuration have to be adjusted accordingly (e.g. as indicated by 'bin' in Table 2.2, via the Bonferroni correction $p_\alpha' = p_\alpha/n$; cf. Bortz 2005: 129; Gries 2008: 244; Sigley 2003). If a configuration then turns out to be significantly more frequent than expected ('Obs-exp' = '>') it is called a 'type', and if it is less frequent than expected ('Obs-exp' = '<') it is said to be an 'antitype' (cf. Gries 2008: 245–6). Since the *p*-values of configurations can be fairly small (in Table 2.2 2.898e-06 = 0.000002898), the significance value is furthermore often indicated by a series of asterisks ('Dec' in Table 2.2, with '*' ≅ $p<0.05$, '**' ≅ $p<0.01$ and '***' ≅ $p < 0.001$). For the sake of readability, however, in this book I will only give the significance thresholds in the CFA tables (i.e. $p <0.05$, $p<0.01$, <0.001), with the precise *p*-value figures being

Table 2.2 *Fictitious output of a CFA*

	FACTOR_1	[…]	FACTOR_n	Frequ	Exp	Cont.chisq	Obs-exp	P.adj.bin	Dec	Q
1	BritishE	.	passives	621	709.5	11.039	<	2.898e-06	***	0.125
2	KenyanE	.	passives	798	709.5	11.039	>	2.898e-06	***	0.125
[…]	[…]	[…]	[…]	[…]	[…]	[…]	[…]	[…]	[…]	[…]

made available in the Appendix for each CFA. Finally, *p*-values are affected by sample size and therefore CFAs also include a parameter called 'coefficient of pronouncedness' ('Q'), which is a sample size-independent measure of effect size ranging from 0 to 1 (thus equivalent to r^2; cf. Bortz, Lienert and Boehnke 1990: 156; Gries 2008: 252).

If CFAs employ chi-square tests, however, they obviously require the expected frequencies of configurations to be greater than 5 (see above; also Woods, Fletcher and Hughes 1986: 144f.). In cases where this criterion cannot be met, as will often be the case in linguistic studies, the significance of configurations must be calculated by the exact binomial test (cf. Bortz, Lienert and Boehnke 1990: 157). Due to the fact that the *HCFA 3.2* script (Gries 2004a) for R automatically carries out an exact binomial test, it was decided to use this program for the present study. Another advantage of the *HCFA 3.2* script was that it does not run a simple CFA but a hierarchical CFA (thus the name *HCFA*) over the data. While CFAs only test the effect of a set of specified configurations (say, FACTOR 1 × FACTOR 2 × FACTOR 3), HCFAs generally also check whether there are effects of simpler configurations (e.g. FACTOR 1 alone or FACTOR 1 × FACTOR 2; cf. Gries 2008: 249–54).

After having presented the first type of evidence used in the present study together with the statistical tools for its interpretation, next I will focus on the second, corroborating type: experimental introspection data.

2.3 Exhibit B: Introspection data

Following Schütze (1996) and Cowart (1997), it was decided that introspection data can only be considered valid and reliable if they are collected and interpreted in a scientifically sound way, following the design of psycholinguistic experiments.

2.3.1 Experiment design

A first important prerequisite for the scientific design of an experiment is that the materials used have been created with the help of so-called 'paradigmlike token sets' (Cowart 1997: 13). The underlying idea behind this approach is that it is well known that '… an informant's response to an individual sentence may be affected by many different lexical, syntactic, semantic, and pragmatic factors, together with an assortment of extralinguistic influences that become haphazardly associated with linguistic materials and structures' (Cowart 1997: 46). In order to minimize these confounding factors, paradigmlike token sets ensure that all these factors are uniformly spread across all the tested items and that differences of judgements of two items can thus solely be attributed to the syntactic phenomenon under investigation. The first step to achieve this goal is to take a particular lexicalization of

Table 2.3 *Token set example* laugh at: You wouldn't believe the things …

Token set	Pied-piping	Stranded
which	… at which Bill laughs	… which Bill laughs at
that	… at that Bill laughs	… that Bill laughs at
Ø	… at Ø Bill laughs	… Bill laughs at

a phenomenon and cross all tested conditions until all theoretically possible variants have been created.

Take, for example, an experiment testing preposition placement with prepositional verbs in English relative clauses. Since there are two variants of PREPOSITION PLACEMENT (stranded vs pied-piped) and three types of RELATIVIZER (*wh-* vs *that* vs Ø), the resulting token set contains six possible structures (2 types of PREPOSITIONAL PLACEMENT × 3 types of RELATIVIZER; see Cowart 1997: 46–50). Table 2.3 illustrates all these six structures for the prepositional verb *laugh at* in the sentence *You wouldn't believe __ the things Bill laughs __*.

Nevertheless, before running an experiment there is still no way to preclude the possibility of particular lexical effects. So if only the six sentences from Table 2.3 were presented to a single subject, the results might be skewed, for example if for some idiosyncratic reason speakers favoured pied-piping with *laugh at* but not with other prepositional verbs. In addition to this, it is well known that each confrontation with a particular stimulus changes an informant, which might also influence his/her judgement. For these reasons all experimental stimuli must also be 'counterbalanced' (i.e. every subject sees all conditions, but never with the same lexical material; Cowart 1997: 93). Thus token sets with six different lexicalizations (e.g. *apologise for*, *dream of*, *rely on*, *sleep with*, *talk about*) for every PREPOSITION PLACEMENT × RELATIVIZER factor combination must be created. Then a different factor combination can be taken from each of the six token sets to generate a set of stimuli which contains all conditions but only one sentence from each token set, yielding a so-called 'material set' such as Table 2.4.

Furthermore, it is necessary that all lexicalizations of a phenomenon be judged by a different subject, and all informants judge all factor combinations but that no informant see more than one sentence from a single token set (Cowart 1997: 93).

While counterbalancing thus avoids many potentially confounding factors, further precautions are still necessary to ensure objective introspection data elicitation. One additional problem that can affect psycholinguistic experiments is the formation of implicit hypotheses. If an informant were only exposed to experimental stimuli he/she might become aware of the aim of the experiment which in turn might affect and distort his/her judgements.

Table 2.4 *Counterbalanced material set sentence list*

Factor combination	Lexicalization
pied-piping + *wh-*	I know the man **on whom** Jane relied.
stranding + *wh-*	Sally fancies the guy **who** Steve talked **about**.
pied-piping + *that*	You wouldn't believe the things **at that** Bill laughs.
stranding + *that*	Brad did something **that** he apologised **for**.
pied-piping + Ø	Sarah never achieved the fame **of** Ø she dreamt.
stranding + Ø	Jennifer never calls the groupies Ø she sleeps **with**.

In order to preclude such effects it is important to include at least as many balanced, i.e. grammatical and ungrammatical, fillers as experimental stimuli. These fillers then act as distractors and prevent informants from forming implicit hypotheses (cf. Cowart 1997: 93).

In addition to this, '[f]atigue, boredom, and response strategies the informant may develop over the course of the experiment can have differing effects on sentences judged at various points in the entire procedure' (Cowart 1997: 94). It is therefore essential to randomize the order in which the stimuli and the fillers are presented to the informant. This is also important since earlier research (cf. Bock 1987, 1990; Cowart 1997: 51–2) has shown that the preceding sentence can influence the judgement of a following sentence. Only the randomized presentation of the experimental items can guarantee that such order effects do not 'systematically distort effects to the targeted differences among sentence types' (Cowart 1997: 51).

As a result of the above considerations, all the experiments presented in this book were designed using token sets which were counterbalanced and supplemented by at least as many fillers and the order of all stimuli was randomized for each informant (for information on the total number of stimuli used per experiment see chapter 5). Finally, since the corpus data for British English as well as Kenyan English come from speakers from the upper end of the sociolinguistic continuum, the informants recruited were all adults (over 18) who have at least completed secondary education (for details of the sampled subjects, see chapter 5).

2.3.2 *Methodology: Magnitude Estimation*

After having outlined the design of the experimental stimuli, I now want to give an overview of the specific methodology employed to elicit introspection data. The method used in the experiment was based on the experimental paradigm of Magnitude Estimation (see Bard, Robertson and Sorace 1996; Keller 2000). As psychophysical experiments have shown, human beings are not really good at making absolute judgements, e.g. saying whether a line is 10 or 15 cm long. Instead, in Magnitude Estimation studies subjects are asked to judge stimuli relative to a reference item. Thus subjects judge whether a given line is longer or shorter than a reference line and try to express this difference in numerical terms (e.g. saying that a stimulus is half as short as the reference line). Since such relative judgements seem easier for humans to make, this approach allows gathering of far more reliable results. Recently, several studies (e.g. Bard, Robertson and Sorace 1996; Featherston 2004, 2005; Keller 2000; Keller and Alexopoulou 2005) have applied this methodology to sentence-judgement experiments. Thus they asked subjects to give numerical judgements on sentences proportional to a constant reference sentence. The results from these Magnitude Estimation studies indicate that eliciting linguistic judgements via this method allows 'reliable and fine-grained measurements of linguistic intuitions' (Keller and Alexopoulou 2005: 1120). The experiments for this study were all conducted using the WebExp software (Keller *et al.* 1998), which includes a cross-modality (judgement of line length) as well as a linguistic training session and automatically randomizes the order of presentation of stimuli in the main experiment (for more information on WebExp, see Keller 2000; Keller and Alexopoulou 2005).

The WebExp software allows running Java-based on-line acceptability experiments as well as creating printed versions of these experiments (Keller *et al.* 1998: 7, 12). The greatest advantages of online experiments are clearly that subjects cannot go back and change earlier answers and that they must respond to all experimental items (since otherwise the software will not allow them to proceed). On the other hand, notable disadvantages are that subjects are self-selecting (i.e. that only a limited part of a speech community uses the internet and is willing to participate in such studies) and that subject authentication is not as good as under laboratory conditions (Keller *et al.* 1998: 6). Furthermore, it turned out that due to technical problems and limited internet access no online data could be gathered from the Kenyan speakers at all.

For the present study it was decided to use the online version of an experiment whenever possible since the advantages clearly outweigh the disadvantages (especially since the self-selecting subjects willing to participate in such studies turn out to be members of the targeted population: educated adults). Whenever this was not possible – either due to technological problems or low response rates – the experimenter personally distributed the printed questionnaires to a set of subjects, explicitly informing them about the restrictions of the experiment (i.e. not to spend more than a few seconds on a single item, not to go back and change earlier answers and to check that they have judged all items).

After this short introduction to the WebExp experiments, the reader might wonder whether Magnitude Estimation is in fact a valid way of eliciting linguistic introspection data. On the one hand, extending the Magnitude Estimation method to linguistic introspection experiments might seem straightforward: since 'linguistics is a branch of psychology that studies a specialized kind of human perception, it is a sister field to psychophysics' (Bard, Robertson and Sorace 1996: 38). On the other hand, ever since Chomsky's *Syntactic Structures* much of mainstream linguistics has taken it for granted that '[t]he fundamental aim in the linguistic analysis of a language L is to separate the grammatical sequences which are the sentences of L from the ungrammatical sequences which are not sentences of L and to study the structure of the grammatical ones' (Chomsky 1957: 13). However, if sentences can simply be divided into either grammatical or ungrammatical ones, then the question arises why one should carry out Magnitude Estimation introspection experiments which allow subjects to make far more intermediate judgements?

One obvious reason is that introspection judgements are not judgements on the grammaticality of a structure as such but on its acceptability. Grammaticality only pertains to the ontological status of a linguistic stimulus, i.e. in a technical sense, whether a string is generated by a grammar or not. An acceptability judgement, on the other hand, is how a speaker rates a linguistic stimulus (see Bard, Robertson and Sorace 1996: 33). The acceptability of a sentence is considered to depend not only on its grammaticality but also its naturalness in discourse or its immediate comprehensibility (see Chomsky 1965: 10). Furthermore, unlike grammaticality, 'acceptability will be a matter of degree' (Chomsky 1965: 10), i.e. that '[t]he more acceptable sentences are those that are more likely to be produced, more easily understood, less clumsy and in some sense more natural' (Chomsky 1965: 11).

From the description of the design of the introspection experiments in section 2.3.1 it should have become apparent, however, that the aim of the present study was not to establish the discourse naturalness of a particular preposition placement construction. Indeed, the employment of counterbalanced token sets was used explicitly to relate different acceptability judgements to grammatical phenomena (e.g. preposition-stranding with different relativizers; cf. section 2.3.1). Since grammaticality does at least play some part in the acceptability judgement of a sentence, statistical differences in judgements of conditions of counterbalanced token sets can therefore not simply be explained away as merely the effect of discourse naturalness.

Interestingly, some support for fine-grained acceptability studies as a window on fine-grained grammaticality differences even comes from mainstream generative linguists: already in the 1960s Chomsky recognized, for example, that there are 'degrees of grammaticalness' (1965: 148). While he still maintained that the class of grammatical sentences constituted a single set, he identified various different degrees of ungrammaticality (e.g. distinguishing subcategorization violations such as *John found sad* from selectional rule violations such as *Colorless green ideas sleep furiously*; see Chomsky 1965: 148–53). Even in a non-technical sense the idea that there are degrees of ungrammaticality can be made intuitive; cf. e.g. the following sentences (2.5a–c):

(2.5) a. They are killing the ducks.
b. They is killing the ducks.
c. They is kill the ducks.

While (2.5a) is grammatical, in Standard English the string in (2.5b) is ungrammatical since it contains an agreement error (*they is*). In addition to this agreement violation, (2.5c) also lacks the required *-ing* participle marker on the verb *kill*. Accordingly, (2.5c) can be said to be more ungrammatical than (2.5b) (for a more technical explanation, see Chomsky 1961).

Since violations of grammaticality are cumulative and also of differing ontological status, it can be taken for granted that there are at least degrees of ungrammaticality. In order to properly investigate such degrees of ungrammaticality it takes a fine-grained and carefully planned introspection experiment as described in section 2.3.1. In fact, as Sorace and Keller have shown, Magnitude Estimation experiments are particularly well suited for distinguishing, for example, semantic from syntactic violations (or in their terminology 'soft' and 'hard constraint violations'; Sorace and Keller 2005: 1502–3): while the former only result in slightly lower acceptability scores, the latter lead to significantly decreased scores.

Moreover, despite the fact that it is still widely assumed that grammatical sentences are a uniform categorical category, there is an increasing number of linguists who contest this view. Ever since the advent of prototype theory (Rosch 1978; Wittgenstein 1953) it has become clear that speakers often conceptualize linguistic categories not as clear-cut but as fuzzy, with some entities being more central and others more peripheral (see also e.g. Labov 1973; Lakoff 1987; Taylor 1995; and for an excellent reader on the topic of linguistic categorization, B. Aarts *et al.* 2004). As mentioned in chapter 1, there is a growing body of work within usage-based construction grammars that considers language acquisition an input-driven bottom-up process (cf. Croft 2001; Croft and Cruse 2004; Langacker 2005; Tomasello 2003; and chapter 6). In such approaches more frequent exemplars of a syntactic entity are also more likely to become more deeply entrenched cognitively, thus qualifying as better exemplars of the emerging prototypical mental concept. Drawing on an empirical investigation of particle movement in English, Gries has furthermore argued that better exemplars of a prototypical concept can also be expected to receive higher acceptability judgements than less typical ones (2003: 132–9). Accordingly, statistically significant differences between conditions in experiments employing counterbalanced token sets can be said to be indicative of different degrees of cognitive entrenchment, or in other words of prototypicality effects.

Magnitude Estimation experiments allow speakers to differentiate as many intermediate levels of acceptability as they deem necessary. Yet, if statistically significant differences in judgements can either be interpreted as differences in the degree of entrenchment of grammatical constructions or the degree of ungrammaticality, then how can these two be distinguished? As I have argued elsewhere (Hoffmann 2006), it is important in such cases to contrast the judgement scores of the experimental items with those of the fillers. Since all judgements are relative within Magnitude Estimation experiments the set of grammatical and ungrammatical fillers constitute the background against which the effects of the experimental stimuli need to be interpreted. Besides, as mentioned above, the use of a corroborating source of evidence, i.e. corpus data, is of paramount importance.

2.3.3 Forensic tools II: SPSS

All Magnitude Estimation data elicited for the present study obviously had to be 'forensically' examined, i.e. subjected to a statistical analysis. For this, the data were entered into the SPSS 12.0 for Windows program. Due to the fact that subjects had employed different scales and values for their judgements, the data were then first normalized by transformation to *z*-scores (Featherston 2004, 2005). This procedure 'effectively unifies the different scales that the individual subjects adopted for themselves, and allows [one] to inspect the results visually' (Featherston 2005: 1533).

Magnitude Estimation yields response variables which are measured on an interval scale (Bard, Robertson and Sorace 1996: 39), while the experiments presented in this book only tested categorical explanatory variables. In addition to this, all experiments crossed several factors so that more than two means had to be compared. Furthermore, as a result of counterbalancing the token sets (see section 2.3.2) all subjects were tested on all conditions of the experiment. Such a so-called within-subject design is said to use repeated measures and the scores it yields had to be analysed via a 'repeated measure Analysis of Variance' (ANOVA; for details see Bortz 2005: 331–60; Crawley 2005: 154; Field 2003, 2009: 457–505; Gravetter and Wallnau 1992: 293).

Repeated measure ANOVAs require that the differences between all pairs of treatment levels have approximately equal variances (Field 2009: 459). If this assumption, known as 'sphericity', is not met, the statistical computation of a repeated measure ANOVA has to be corrected accordingly. In SPSS the so-called 'Mauchly's test of sphericity' tests the data for this assumption. If the Mauchly's test yields a significant *p*-value (below 0.05), the results of the following repeated measure ANOVA have to be corrected using an appropriate correction. SPSS offers several options in such cases, but for the present study the conservative Huynh and Feld correction was always applied (see Field 2009: 460–1).

In ANOVA tests, the critical value for identifying significant effects is the *F*-ratio calculated by dividing the differences between treatments by the differences within treatments (see e.g. Gravetter and Wallnau 1992: A-90). For each experiment two repeated measure ANOVAs had to be carried out: one testing whether the effects are significant by subject (F_1), and a second one to see whether they were also significant by item (F_2). This was necessary since

> just as the informants actually tested in an experiment are (usually) seen as representatives of the entire population from which they are selected, the token sets are likewise seen as representatives of all the relevantly similar token sets one might construct in the same language (or perhaps any language). Just as statistical tests on data for informants test the reliability of patterns seen in those results, so tests on data for token sets test the reliability of patterns seen in the summaries of the token set data. (Cowart 1997: 122)

Once the *F*-ratio score of an explanatory variable (also called 'factor' in ANOVAs) was identified as significant in either the by-item or the by-subject analysis, the next question to be addressed concerned the informativeness of this result: i.e. how much variation in the data can actually be explained by a significant factor? The statistical value which provides this information for ANOVAs is the so-called eta-square parameter η^2. This value is calculated by dividing the '[s]um of squares (SS) for each factor or interaction in the ANOVA result table … by the total sum of squares for the set as a whole' (Cowart 1997: 136). In non-technical terms, η^2 can therefore be seen as the proportion of data explained by a particular factor.

Note that a significant ANOVA result only indicates that there is an effect within a factor. As long as there are only two levels or conditions within a factor this does not cause any problems since a straightforward comparison of the means of the two levels will immediately allow identification of the effect. Whenever there are three or more levels, however, the locus of the significant effect cannot be found that easily. Instead, the standard procedure in such cases is to run a post-hoc test over the ANOVA model. Such post-hoc tests involve a pairwise comparison of all levels which is corrected for multiple testing (Gravetter and Wallnau 1992: 372). For the present study the Tukey post-hoc test (Bortz 2005: 325–8; Gravetter and Wallnau 1992: 406) was used whenever a significant factor had more than two levels.

As argued in section 2.3.2, in order to decide whether a particular effect should be regarded as reflecting the grammaticality or the ungrammaticality of a condition it was important to compare these with the set of grammatical and ungrammatical filler items. Because the fillers were not part of the repeated measure ANOVAs, such comparisons had to be carried out by performing a set of dependent *t*-tests.[11] To preclude an inflation of the α-error in these tests the *p*-values for multiple *t*-tests were adjusted using

[11] 'Dependent' here has the same meaning as 'repeated measures' in the ANOVA: since all subjects judge the same experimental conditions and the same set of fillers, the two treatment conditions tested do not qualify as independent.

the standard, albeit rather conservative, Bonferroni-correction $p_{\alpha}' = p_{\alpha}/n$. (see Bortz 2005: 129; Sigley 2003; and section 2.2.2).

Finally, in order to visualize the results from the statistical analysis of the introspection data, graphs displaying the mean judgements of significant results together with standard error bars were created using Excel for Windows.

In this chapter I have argued that linguists should treat the analysis of a particular syntactic phenomenon as a criminal case which draws on corroborating evidence from corpora and introspection. In addition to this, I have laid open the kind of evidence as well as the forensic tools (a.k.a. statistical programs) used for the present case. Before turning to an in-depth forensic analysis of the evidence, however, it is important to first survey witness statements and the case notes of other detectives. In other words, the discussion of preposition placement in the linguistic literature has to be examined in the next chapters.

3 Case notes: Independent factors

After having presented the book's methodological and statistical background, it now becomes necessary to see which factors have so far been claimed to influence preposition placement in English. As will be seen, these include clause type (3.1), the type of PP (3.2), the level of formality (3.3), the phrase in which the PP is contained (3.4) and processing complexity (3.5). All of these might be effective in both British and Kenyan English. However, as section 3.6 will show, there might also be second-language-specific factors at work in Kenyan English.

3.1 Construction-specific constraints

As Pullum and Huddleston point out (cf. 2002: 627), in English the following four structures allow a choice between preposition-stranding and pied-piping:

(3.1)	a.	**[Stranding]**$_i$ I've heard **of**$_i$.	[preposing]
	b.	**[What]**$_i$ is he talking **about**$_i$?	[interrogative]
	c.	**[What a great topic]**$_i$ he talked **about**$_i$!	[exclamative]
	d.	the structure [**[which]**$_i$ he talked **about**$_i$].	[*wh*-relative]
(3.2)	a.	**[Of stranding]**$_i$ I've heard$_i$.	[preposing]
	b.	**[About what]**$_i$ is he talking$_i$?	[interrogative]
	c.	**[About what a great topic]**$_i$ he talked$_i$!	[exclamative]
	d.	the structure [**[about which]**$_i$ he talked$_i$].	[*wh*-relative]

The above examples illustrate that variable preposition placement in English is possible in topicalized ((3.1a, 3.2a); 'preposed' in Pullum and Huddleston's terms), interrogative (3.1b, 3.2b), exclamative (3.1c, 3.2c) and *wh*-relative clauses (3.1d, 3.2d). In addition to this, however, there are also syntactic environments in which only stranding is permitted:

(3.3)	a.	the structure [(**that**)$_i$ he talked **about**$_i$].	[non-*wh*-relative]
	b.	**the same stuff**$_i$ as [I talked **about**$_i$].	[comparative]
	c.	**His thesis**$_i$ was easy [to find fault **with**$_i$].	[hollow]
	d.	**Stranding**$_i$ has been talked **about**$_i$ enough.	[passive]

Non-*wh*-relative clauses (3.3a), comparative (3.3b), hollow ('non-finite clauses … other than relatives or open interrogatives where some non-subject NP is missing but recoverable from an antecedent NP or nominal'; Huddleston 2002c: 1245; (3.3c)) and passive clauses (3.3d) thus allow preposition-stranding, but not pied-piping:

(3.4) a. *the structure [[**about (that)**]$_i$ he talked$_i$]. [non-*wh*-relative]
b. ***the same stuff**$_i$ as [**about** I talked$_i$]. [comparative]
c. ***His thesis**$_i$ was easy [**with** to find fault$_i$]. [hollow]
d. ***About stranding**$_i$ has been talked$_i$ enough. [passive]

As pointed out in chapter 2, preposition placement can be analysed as a linguistic variable ('dependent variable'). The above examples show that the factor group CLAUSE TYPE contains a number of categorical (the constructions which obligatorily demand stranding in (3.3)) as well as variable (the constructions in (3.1/3.2)) independent factors that affect this dependent variable. In addition to this, however, a comparison of (3.2a) and (3.4a) also reveals an interaction effect: depending on the relativizer, relative clauses either license both stranding and pied-piping (if a *wh*-relativizer is present; see (3.1a/3.2a)) or obligatorily lead to stranding (if a *that*- or Ø- (i.e. no overt) relativizer is used; cf. (3.3a/3.4a)). The complexity of this situation warrants a closer look at preposition placement in relative clauses.

3.1.1 Relative clauses

(3.5) I know **the girl**$_i$ **who**$_i$ **John went to school with** $\boldsymbol{\Delta}_i$

The bold-faced expression in (3.5) is a typical example of a so-called '(bound) relative construction': the noun *girl* ('the antecedent or head noun') has a clausal postmodifier ('the relative clause') which is introduced by the 'relativizer' *who*. As the co-indexation in (3.5) shows, the relativizer has the function of linking the antecedent to a gap position 'Δ_i' in the relative clause. The intuition behind this co-indexation is based on the fact that logically *girl* has two functions: it is an argument of *know* in the matrix clause and at the same time argument of *with* in the relative clause. Example (3.5) could therefore be paraphrased as 'I know a girl and John went to school with this girl' (see Biber *et al.* 1999: 608; Huddleston, Pullum and Peterson 2002: 1034; Olofsson 1981: 18ff.; Schmied 1991a: 17ff.).

As the sentences in (3.6) illustrate, Standard British English has a number of lexemes functioning as relativizers in finite relative clauses:

(3.6) a. *the day **on which*** she arrived / *the day **which*** she arrived ***on***
b. *the day **that*** she arrived ***on***
c. *the day **Ø*** she arrived ***on***
d. *the day **when*** she arrived

e. *the day* **Ø** she arrived
f. *the day* ***that*** she arrived (taken from Quirk *et al.* 1985: 1254)

Biber *et al.* (1999: 608) classify relativizers such as *that*, *Ø* 'ZERO' (i.e. the phonologically empty form in (3.6c)) and the *wh*-forms *which*, *who*, *whom*, *whose* as 'relative pronouns'. The *wh*-lexemes *where*, *when* and *why*, on the other hand, are regarded as 'relative adverbs' (Biber *et al.* 1999: 608), since their distribution is limited to relative clauses such as (3.6) in which the gap position is an 'adverbial expression of place, time, and cause' (Quirk *et. al* 1985: 1253).

Examining the sentences in (3.6), the first interesting observation concerning preposition placement about British English relative clauses comes from a comparison of examples (a) and (d): relative clauses with *wh*- pronouns and adverbial gaps governed by a preposition (e.g. *on which/which … on*) compete with an alternative in which a relative adverb replaces both, *wh*-pronoun and preposition (here *when*; cf. Biber *et al.* 1999: 624). Interestingly, although Biber *et al.* do not give *that* and *Ø* as possible relative adverbs, these also appear in constructions without the adverbial preposition (3.6e, f). The *wh*-relative pronouns, on the other hand, cannot appear without the preposition:

(3.6) g. *the day **which** she arrived

Furthermore, as pointed out above, if the preposition is overtly realized, its placement is restricted by the relativizer choice: whereas *wh*-relative pronouns allow both preposition-stranding and pied-piping (see (3.6a)), prepositions always have to be stranded in *that*- and *Ø*-introduced relative clauses (cf. (3.6b, c) vs (3.7a, b)):

(3.7) a. *the day **on that** she arrived
b. *the day **on Ø** she arrived

Non-finite relative clauses impose even stronger restrictions on the relativizer choice and preposition placement than finite ones: first of all, non-finite relative clauses can be introduced by *Ø* (again with obligatory preposition-stranding, cf. (3.8a)), but do not allow a *that*-relativizer (3.8b); Huddleston, Pullum and Peterson 2002: 1057):

(3.8) a. a day **Ø** to arrive **on**
b. * a day **that** to arrive **on**

In addition to the ban on *that*, non-finite relative clauses exhibit an interesting categorical effect on preposition placement if introduced by a *wh*-pronoun: whereas *wh*-pronouns in finite relative clauses permit both stranding and pied-piping (see (3.6a)), they cannot occur with stranded prepositions in non-finite clauses (Hoffmann 2005: 263; Huddleston, Pullum and Peterson 2002: 1067):

(3.9) a. a day **on which** to arrive
b. *a day **which** to arrive **on**

This interaction effect of PREPOSITION PLACEMENT × CLAUSE TYPE × *WH*-PRONOUN is particularly remarkable since non-finite interrogative clauses allow both stranding and pied-piping with *wh*-pronouns (Sag 1997: 462):

(3.10) a. I don't know **on which** day to arrive
b. I don't know **which** day to arrive **on**

Due to the fact that non-finite relative clauses demand an overt *wh*- relativizer to pied-pipe the preposition, van der Auwera (1985: 166) claims that the ungrammaticality of (3.9b) basically accounts for the ungrammaticality of (3.8b): since *that* can never pied-pipe a preposition, it cannot be used in non-finite relative clauses, where pied-piping is obligatory. Such reasoning, however, still leaves two questions unanswered: (1) why is it impossible to pied-pipe prepositions in Ø- and *that*-introduced relative clauses, and (2) why is pied-piping obligatory in non-finite relative clauses with overt *wh*-relativizers, while both stranding and pied-piping are possible in non-finite *wh*-interrogative clauses?

In addition to the above factors, the distribution of preposition placement is further complicated by the existence of a special type of relative clauses in which a *wh*-word acts simultaneously as antecedent and relativizer, the so-called 'fused' or 'free relative clauses':

(3.11) a. What he is talking about is called feminism
b. That which he is talking about is called feminism

As the paraphrase in (3.11b) shows, the *wh*-word *what* functions both as the antecedent (cf. the demonstrative pronoun *that* in (3.11b)) and the relativizer (cf. the relative pronoun *which* in (3.11b)) of the relative clause in (3.11a). Due to this, free relative clauses are generally considered to disallow pied-piping (3.12a) since putting the preposition before the *wh*-item in these cases would mean that it would precede not only the relativizer but also the antecedent ((3.12b); cf. Pullum and Huddleston 2002: 628–9). Consequently, the preposition in (3.12a) would be extracted out of the relative clause into the matrix clause, rendering the resulting structure ungrammatical:

(3.12) a. *About what he is talking is called feminism
b. *About that which he is talking is called feminism

Finally, there is a subtype of relative clauses that does not appear to exert any categorical effect on preposition placement but is interesting nevertheless since it seems to have three competing variants: relative clauses in *it*-cleft-sentences (Biber 1999: 959; Ward, Birner and Huddleston 2002: 1416–18).

(3.13) a. It was John **who** I talked **to**
b. It was John **to whom** I talked
c. It was **to John** that I talked

It-clefts such as (3.13) always consist of *it*, a form of *to be* and a foregrounded/highlighted element (*John* in (3.13a,b) and *to John* in (3.13c)) that is followed by a relative-clause-like element (Biber 1999: 959). As Ward, Birner and Huddleston point out (2002: 1418) the most frequently foregrounded elements in *it*-clefts are NPs (3.13a, b) and PPs (3.14c). Moreover, they show that if the foregrounded element is the complement of a preposition then the preposition can be stranded (3.13a) or in formal style pied-piped to the front of the relative clause (3.13b). Yet, while Ward, Birner and Huddleston discuss the possibility of foregrounding PPs, they do not seem to consider such cases instances of pied-piping. If, however, there is a context in which a complement PP is focused (A: *What did you just say? You talked about John?* B: *No, it was* ***to John*** *that I talked!*), then pied-piping appears to be possible beyond the relative clause (cf. (3.13c); Paul Livesey, pc).

All in all, preposition placement in English can thus be said to be restricted by a great number of categorical and variable effects of the factor group CLAUSE TYPE. Figure 3.1 gives a complete overview of these factors, indicating categorical stranding contexts by white font and categorical pied-piping contexts by light grey font, with clause types allowing variable preposition placement given in a neutral black font.

The complex situation summarized in Figure 3.1 obviously required that the ICE-GB and ICE-EA corpus data were coded for all the CLAUSE TYPE factors given in the figure. Figure 3.1 is thus also a summary of the first factor group of the multivariate corpus analysis. In addition to this, however, the figure also raises a couple of interesting questions.

First of all, the complex interaction effect in relative clauses means that the choice of relativizer will indirectly affect preposition placement in relative clauses. If, for example, a factor favours a *that*- or Ø-relativizer it will consequently also favour stranding whenever the complement position of a prepositional phrase is relativized. It is therefore important to discern between effects which directly affect preposition placement and those which only indirectly influence it (due to their effect on the choice of relativizer). Such situations then beg the fundamental question whether a factor can only affect the choice of relativizer and its effect on preposition placement can be treated as a mere epiphenomenon. Alternatively, it might be conceivable that due to permanent usage such a factor will also become associated as directly favouring stranding. It remains to be seen whether this question can be decided on anything but theoretical grounds (e.g. by resorting to Occam's

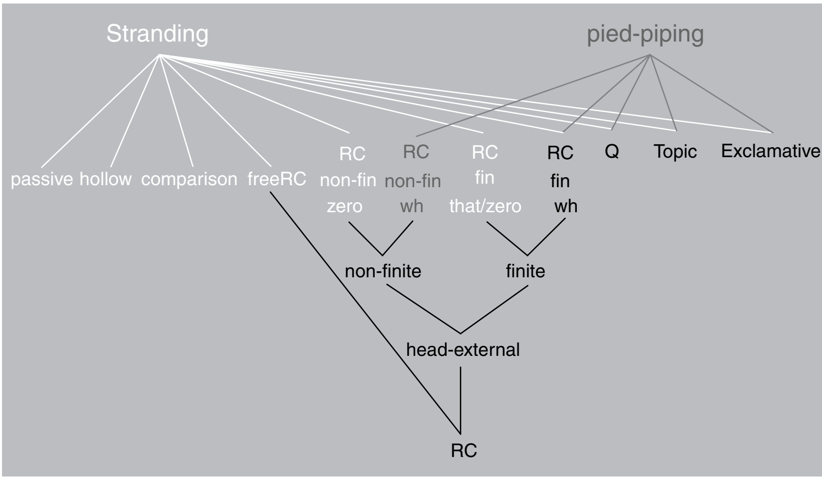

Figure 3.1 Preposition placement across clause types

Razor, i.e. assuming that the single effect on the choice of relativizer alone suffices as an explanation for the data). In any case, it will become necessary to take a closer look at the displaced prepositional complement in the various clause types given in Figure 3.1 in order to see whether there are any further interaction effects. Due to the potential indirect effect in relative clauses pointed out above, this discussion will focus especially on factors influencing the choice of relativizer.

Another important issue concerns the homogeneity of those clause types which seem to exhibit similar effects on preposition placement. For while, for example, passive clauses and free relative clauses are categorical stranding environments it is not guaranteed that both clause types license the same types of stranded prepositions. In addition to this, it would, of course, be theoretically elegant if preposition placement in *wh*-relative clauses and *wh*-interrogative clauses could be captured by a single constraint. It is, however, also possible that both contexts exhibit different preposition placement preferences.

In the following I will first investigate the various types of displaced elements, i.e. the adjacent complements of a pied-piped preposition and the logical complement of a stranded preposition, before turning to potentially idiosyncratic effects of the individual clause types.

3.1.2 Type of displaced element

3.1.2.1 Beyond bound relative clauses

Apart from non-*wh*-relative clauses, which have been dealt with above in section 3.1.1, the remaining categorical stranding clause types already display a great heterogeneity with respect to the displaced element that logically functions as the complement of the stranded preposition:

(3.14)	a.	[**Pied-piping**]$_i$ has been talked **about**$_i$ enough.	[passive]
	b.	He talked about **the same stuff**$_i$ as [I talked **about**$_i$].	[comparative]
	c.	**His ideas**$_i$ were easy [to find fault **with**$_i$].	[hollow]
	d.	**What**$_i$ he is talking **about**$_i$ is called stranding.	[free relative]

As can be seen in (3.14), the displaced element in passives (*pied-piping* in (3.14a)), comparatives (*the same stuff* in (3.14b)) and hollow clauses (*his ideas* in (3.14c)) is usually a full NP. In contrast to this, the logical complement of the stranded preposition in free relative clauses is a *wh*-word (*what* in (3.14d)) that can often carry an *-ever* suffix (e.g. *Whatever he is talking about usually has something to do with stranding*). The important observation here is that the 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. This claim can be further corroborated by looking at the clauses that allow both stranding and pied-piping (for *wh*-bound relative clauses see section 3.1.1 above):

(3.15)	a.	**[Stranding]**$_i$ I've heard **of**$_i$ enough.	[preposing]
	b.	**[Of stranding]**$_i$ I've heard$_i$ enough.	[preposing]
(3.16)	a.	**[What]**$_i$ is he talking **about**$_i$?	[interrogative]
	b.	**[About what]**$_i$ is he talking$_i$?	[interrogative]
(3.17)	a.	**[What a strange topic]**$_i$ he talked **about**$_i$!	[exclamative]
	b.	**[About what a strange topic]**$_i$ he talked$_i$!	[exclamative]

Thus, while displaced NP elements obligatorily co-occur with stranded prepositions in passives, comparatives and hollow clauses, they can pied-pipe or strand prepositions in preposed/topicalized structures (cf. (3.15)). Moreover, when it functions as the displaced element in free relative clauses the *wh*-word *what* is always the logical complement of a stranded preposition, but in interrogatives or exclamatives it can be preceded by a pied-piped preposition (see (3.16b) and (3.17b), respectively) or followed by a stranded preposition ((3.16a) and (3.17a), respectively).

The only pair of displaced elements which clearly seem to affect preposition placement are *who* and *whom*:

(3.18)	a.	**[Who]**$_i$ did he talk **about**$_i$?	[interrogative]
	b.	**[Whom]**$_i$ did he talk **about**$_i$?	[interrogative]
(3.19)	a.	the person [**[who]**$_i$ he talked **about**$_i$]	[*wh*-relative]
	b.	the person [**[whom]**$_i$ he talked **about**$_i$]	[*wh*-relative]

Both *wh*-words can co-occur with stranded prepositions in interrogative and bound *wh*-relative clauses ((3.18) and (3.19), respectively; cf. Payne and

Huddleston 2002: 464–6).[1] However, in pied-piped structures *whom* 'is normally the only option' (Payne and Huddleston 2002: 465):

(3.20) a. *[**About who**]$_i$ did he talk$_i$? [interrogative]
b. [**About whom**]$_i$ did he talk$_i$? [interrogative]
(3.21) a. *the person [[**about who**]$_i$ he talked$_i$] [*wh*-relative]
b. the person [[**about whom**]$_i$ he talked$_i$] [*wh*-relative]

As the asterisk '*' in (3.20a) and (3.21a) indicates, pied-piping with *who* is generally considered ungrammatical in English (cf. e.g. van der Auwera 1985; but see below for potential exceptions). Although present-day English hardly has any case-marking, the *who–whom* pair is taken as one of the last remnants of the earlier synthetic system. The ungrammaticality of (3.20a) and (3.21a) is therefore usually attributed to case misassignment: under this view, a preceding pied-piped preposition obligatorily requires the following *wh*-element to be marked for oblique case (or as Payne and Huddleston 2002 or Sag 1997 call it accusative case). Since *who* is said to carry nominative case (according to e.g. Payne and Huddleston 2002 or Sag 1997) its case feature will clash with that required by a pied-piped preposition.

Note that *who* is normally associated with informal and *whom* with formal style (Payne and Huddleston 2002: 464), and that preposition-stranding is often considered more typical of informal style than pied-piping (see further section 3.3). These facts, together with data like (3.20a) and (3.21a) have given rise to an attempt by Radford (1997) to account for all contexts which allow both stranding and pied-piping within a Minimalist Program framework (Chomsky 1995, 2000). His first observation concerns the fact that in most of the world's languages *wh*-words (interrogative and relative pronouns) obligatorily pied-pipe prepositions. German is a case in point (Dekeyser 1990: 94):

(3.22) a. **Über wen** hat er gesprochen?
'**About whom**$_{ACC}$ has he talked?'
b. ***Wen** hat er **über** gesprochen?
Whom$_{ACC}$ has he talked **about**?'
(3.23) a. die Person, **über die** er sprach
'the person **about whom** $_{ACC}$ he spoke'
b. *die Person, **die** er **über** sprach
'the person **whom**$_{ACC}$ he spoke **about**'

[1] Exclamatives only license *how* and *what* as displaced elements (see Huddleston 2002b: 918), free relative clauses only very rarely allow '[*w*]*ho*, *whom*, *whose*, *which*, *why* and *how* [... while] the *-ever* series of forms occur frequently' (Huddleston, Pullum and Peterson 2002: 1071). All the remaining non-relative clause types only have full NPs as displaced elements.

He then assumes that a syntactic case-marking/assignment[2] operation is responsible for the obligatory pied-piping of prepositions in languages like German: according to Radford, the first step of the syntactic derivation of (3.22) and (3.23) is the concatenation or 'merger' of the preposition and the *wh*-word:[3]

(3.24) [über] + $[\text{w-}]_{\text{wh\_Q}}$ → **[über wen]**$_{\text{wh\_Q}}$ [interrogative]
(3.25) [über] + $[\text{d-}]_{\text{wh\_RC}}$ → **[über die]**$_{\text{wh\_RC}}$ [*wh*-relative][4]

In German the preposition *über* assigns accusative case to its complement.[5] Upon merger with *über* the abstract interrogative and relative pronouns *w*- and *d*- thus get instantiated as the accusative forms *wen* and *die*, respectively. A side effect of this case assignment operation is that the so-called *wh*-features which mark *wen* and *die* as interrogative and relative (indicated by 'wh_Q' and 'wh_RC' in (3.24) and (3.25)) get passed on ('percolate') to the entire PP (Radford 1997: 279).

Then the derivations continue until all the remaining lexical items as well as a functional head F have been concatenated:

(3.26) F_{Q} hat er **[über wen]**$_{\text{wh\_Q}}$ gesprochen [interrogative]
(3.27) F_{RC} er **[über die]**$_{\text{wh\_RC}}$ sprach [*wh*-relative]

The head F carries features which mark the clause as interrogative or relative. These features require a phrase with an appropriate *wh*-feature to move to the front of the clause and check them. Remember that the case-checking operation leads to the automatic percolation of the *wh*-features to the entire PPs in (3.24) and (3.25). Consequently the phrases which carry the *wh*-features and which are therefore selected to move to

[2] Unlike earlier approaches, Minimalist analyses actually treat case phenomena as checking operations and not assignment mechanisms (for details see Chomsky 1995, 2000). Nevertheless, for the present purposes the assignment metaphor is probably more accessible for readers unfamiliar with the Minimalist Program and the choice of terminology does not affect the validity of the argument.

[3] I am ignoring several technical details here which are irrelevant for the present discussion, most notably that Chomsky has dispensed with post-narrow-syntax LF operations by assuming cyclic spell-out, i.e. derivation by phase (cf. Chomsky 2001). My critique of Radford is not affected by any modifications he would have to make to get a derivation by phase version of his analysis.

[4] Again (3.24) and (3.25) are actually simplified representations of the actual derivation. In fact the abstract *w*- and *d*- lexical items are abstract phonological, syntactic and semantic feature combinations which enter the derivation fully specified for (e.g.) their case feature. If checking then takes place within narrow syntax, i.e. before (cyclic) spell-out, only *whom* but not *who* will yield an output that converges (cf. Chomsky 1995, 2000). Again, however, these details are immaterial for the present case in point.

[5] Technically, case assignment is in fact a Probe-Goal checking mechanism (cf. Chomsky 2000).

the front of the clause in (3.26) and (3.27) are the full PPs (cf. Radford 1997: 279):

(3.28) **[über wen]$_{wh\_Q}$** F$_{Q}$ hat er gesprochen [interrogative]
(3.29) **[über die]$_{wh\_RC}$** F$_{RC}$ er sprach [*wh*-relative]

Examples (3.28) and (3.29) are thus the underlying syntactic structures of (3.22) and (3.23), and the obligatory pied-piping of prepositions in German is accounted for by the case assignment operation and the ensuing feature percolation mechanism.

In order to explain the more complex situation displayed in English interrogative and relative clauses, Radford next assumes that case assignment in English in formal registers applies just as in German: immediately upon the merger of the preposition the *wh*-feature percolates to the PP (see (3.30a) and (3.31a)). In contrast to this, case-checking in informal registers is postponed until after the syntactic derivation (i.e. takes place at LF), which means that the *wh*-feature does not percolate to the PP (cf. (3.30b) and (3.31b); Radford 1997: 279):

(3.30) a. [about] + [w-]$_{wh\_Q}$ → **[about whom]$_{wh\_Q}$** [interrogative] [formal]
b. [about] + [w-]$_{wh\_Q}$ → **[about [who]$_{wh\_Q}$]** [interrogative] [informal]
(3.31) a. [about] + [w-]$_{wh\_RC}$ → **[about whom]$_{wh\_RC}$** [*wh*-relative] [formal]
b. [about] + [w-]$_{wh\_RC}$ → **[about [who]$_{wh\_RC}$]** [*wh*-relative] [informal]

The different effect of the level of formality then obviously affects preposition placement once the functional head F attracts the *wh*-feature:

(3.32) a. F$_{Q}$ did he talk **[about whom]$_{wh\_Q}$** [interrogative] [formal]
b. F$_{Q}$ did he talk **[about [who]$_{wh\_Q}$]** [interrogative] [informal]
(3.33) a. F$_{RC}$ he talked **[about whom]$_{wh\_RC}$** [*wh*-relative] [formal]
b. F$_{RC}$ he talked **[about [who]$_{wh\_RC}$]** [*wh*-relative] [informal]

In (3.32a) and (3.33a) case-marking has taken place and the entire PP consequently carries the *wh*-feature, leading to preposition pied-piping with *whom*:

(3.34) **[about whom]$_{wh\_Q}$** F$_{Q}$ did he talk [interrogative] [formal]
(3.35) **[about whom]$_{wh\_RC}$** F$_{RC}$ he talked [*wh*-relative] [formal]

In (3.32b) and (3.33b), on the other hand, no case-marking has taken place so that only *who* carries the *wh*-feature and is therefore attracted, leaving the preposition stranded (cf. Radford 1997: 279):

(3.36) **[who]$_{wh\_Q}$** F$_{Q}$ did he talk **[about]** [interrogative] [informal]
(3.37) **[who]$_{wh\_RC}$** F$_{RC}$ he talked **[about]** [*wh*-relative] [informal]

While Radford's analysis might be appealing due to its simplicity and elegance, it also has a number of flaws: first of all, despite the fact that it is a stipulation that partly explains the data, it is by no means clear why case assignment should lead to obligatory feature percolation. Furthermore, *who* and *whom* are the only *wh*-words which exhibit remnants of case-marking. Radford, however, assumes that his analysis also extends to *wh*-words which are not overtly marked for case. Thus despite the fact that there is no visible morphological evidence, he would argue that in *the place in which she died* case assignment has taken place, while in *the place which she died in* no case-marking has occurred. Apart from its theoretical appeal, this lack of case marking on all *wh*-relativizers except *whom* obviously means that Radford's analysis actually receives only very limited support from the data. Finally, an immediate effect predicted by Radford's analysis would be the ungrammaticality of sentences like (3.38):

(3.38)	a.	**Whom** did he talk **about**	[interrogative]
	b.	the man **whom** he was talking **about**	[*wh*-relative]

Since oblique *whom* is considered a reflex of overt case assignment, its *wh*-features should have percolated to the entire PP, causing *about* to pied-pipe along with the relative pronoun to the head of the clause. The stranded preposition should therefore render the sentences in (3.38) ungrammatical. Yet, although speakers might prefer *who* over *whom* in (3.38), Payne and Huddleston point out that such sentences do not seem to be ungrammatical but 'are certainly attested' (2002: 465). On the internet the following attested examples can be found on an informal football fan site and in a letter to the Editor of the Evening Standard:

(3.39) … when Dave asks the assembled '**Whom** did we sign Scholes **from**?' (www.unitedrant.co.uk/archives/2006/01/picture_the_scene.html, accessed 17 August 2007)

(3.40) I cannot begin to imagine the grief of those **whom** David Mellor writes **about** in 'My Friends Betrayed' (6 December). (www.fco.gov.uk/servlet/Front?pagename=OpenMarket/Xcelerate/ShowPage&c=Page&cid=1007029391629&a=KArticle&aid=1134648925991, accessed 17 August 2007)

While case-marking might have played an important role in earlier stages of English (cf. below), the grammaticality of sentences like (3.39) and (3.40) clearly shows that Radford's (1997) analysis does not even adequately describe the distribution of preposition-stranding for *whom*-relative clauses – the only relativizer that still exhibits case morphology. Even Radford himself later (2004: 211–20) abandoned the percolation-induced-by-case-assignment analysis. Following Chomsky (1995: 264), he now argues that in formal registers of English preposition-stranding is simply banned by the so-called 'stranding constraint'. While this claim is also not without problems (for details, cf. section 3.3 on the impact of the level of formality on preposition placement), it should be noted that this analysis implies that the strong preference of *whom* co-occurring with pied-piped prepositions is simply due to the fact that both are favoured by formal contexts.

Interestingly, however, there are two other constructions in which a *who*-relativizer can be preceded by a preposition that can be explained by Radford's 1997 analysis:

(3.41) A: You should speak to someone. B: **To who**?
(3.42) Who gave what **to who**?

Examples (3.41) and (3.42) indicate that if 'the PP stands alone or in post-verbal position [...*w*]*ho* is acceptable in informal style' (Payne and Huddleston 2002: 465–6). Due to the informality of the two constructions, Radford could argue that no case assignment takes place in these cases. Since *who* furthermore remains *in situ* (i.e. does not have to move), these are the only situations in which an unmarked *wh*-word can be preceded by a preposition.

Despite the fact that (3.41) and (3.42) are not really instances of pied-piping (since all PP elements appear in their canonical declarative clause position), it is remarkable that these are undisputedly grammatical structures which allow a preposition to precede a *who*-relativizer. Why then are questions and relative clauses which contain a P + *who* structure considered ungrammatical (cf. (3.20a) and (3.21a), repeated here as (3.43) and (3.44), respectively)?

(3.43) *[**About who**]$_i$ did he talk$_i$? [interrogative]
(3.44) *the person [[**about who**]$_i$ he talked$_i$] [*wh*-relative]

The status of (3.44) is particularly interesting since Sag (1997: 463) argues that in present-day English *that*, just like *who*, is a relative pronoun which carries nominative case. Consequently, the ungrammaticality of pied-piping with *that* (**the person* ***about that*** *he talked*) would be due to case misassignment on a par with the effect of pied-piping with nominative *who*, instead of *whom* (for a similar line of reasoning, cf. van der Auwera 1985). Yet, as the discussion above has shown, a purely case-based analysis of pied-piping and stranding with *who* and *whom* does not seem to allow an adequate description of the data. Besides, in contrast to *who* (cf. (3.41)/(3.42)), there are no structures in English in which a *that*-relativizer is ever preceded by a preposition. As such, it is suspected that the acceptability of pied-piping with *who* is higher than with *that* (simply because, at least locally, i.e. without the wider syntactic context, P + *who* structures are generated and licensed by the grammar, while P + *that* syntagms never occur). Given that none of the constructions in question were to be expected to surface in the corpus data, it was decided to experimentally test the hypothesis that pied-piping with *that* and *who* is equally unacceptable and consequently due to the same case misassignment effect (see sections 5.2 and 5.3).

Finally, it needs to be pointed out that, in addition to *who* and *whom*, the *wh*-words *which*, *whose*, *where*, *when* and *why* are also used in both relative clauses and interrogatives. Moreover, Standard British English also employs *what* and *how* exclusively as interrogative elements (Biber *et al.* 1999: 87; Ungerer *et al.* 1996: 23–4). Note that just as in relative clauses, adverbial interrogatives can replace both *wh*-pronoun and preposition (cf. Ungerer *et al.* 1996: 23):

(3.45) a. **On which day** will he arrive?
b. **When** will he arrive?

As the above discussion of the displaced elements in the various clause types has shown, only the choice of *who* and *whom* in relatives and interrogatives has a direct effect on preposition placement. In the remaining cases, the type of displaced element was taken to be the result of the specific clause type that it occurs in. Nevertheless, for the sake of completeness all tokens from the corpora were coded for the type of displaced element.

3.1.2.2 Bound relative clauses: Secondary effects

In section 3.1.1 it was argued that the complex situation in relative clauses warranted a closer look at the factors that affect the choice of relativizer since these could be said to exert a secondary, indirect influence on preposition placement. In this chapter I will try to give a concise overview of the most important of these potential secondary effects.

Probably the most prominent factor confining the distribution of relativizers in British English relative clauses concerns the 'relative junction' (Olofsson 1981: 18), i.e. the closeness of the semantic antecedent–relative clause relationship: the restrictive/non-restrictive dichotomy. Traditionally, relative clauses have been divided into:

- restrictive relative clauses: i.e. those which are obligatory since they are necessary to identify the reference of the antecedent, and
- non-restrictive relative clauses: i.e. those which are optional since they only add new information without establishing the antecedent's reference (Downing 1978: 379f.; Quirk *et al.* 1985: 1239).

Applying the above definitions to (3.46) shows that the same relative clause can be either restrictive (b) or non-restrictive (a), depending on the previously given information and the number of John's siblings.

(3.46) a. John has a brother, who lives in London.
b. John has a brother who lives in London (and another living in Edinburgh).

In (3.46a) the reference of the *brother* is established, and the relative clause contains only additional information about this brother. On the other hand, if John has two male siblings, then *a brother* on its own will not suffice to identify the reference of this DP. Thus in (3.46b) the relative clause actually provides information which narrows down the set of possible referents of the antecedent DP.

According to Downing (1978: 381) all languages use restrictive relative clauses, whereas '[s]ome languages apparently have no nonrestrictive R[elative]C[lause]'s; in others they are syntactically quite distinct; in others restrictive and nonrestrictive RC's are syntactically indistinguishable' (Downing 1978: 380). As is widely known, English is one of those languages in which the restrictiveness of a relative clause has syntactic reverberations: non-restrictive relative clauses are only marginally introduced by *that* and never by Ø (cf. Huddleston, Pullum and Peterson 2002: 1059):[6]

(3.47) a. John, whom I have just called, told me …
b. *John, Ø I have just called, told me …
c. *John, that I have just called, told me …

In (3.47) the antecedent is a proper noun whose reference is fully established. The relative clause in these examples therefore only contains optional information and is non-restrictive. As (3.47b, c) show, the use of *that* or Ø in such constructions produces an ungrammatical result.

It must be pointed out that Quirk *et al.* emphasize that although '[t]he distinction between restrictive and non-restrictive is valuable, … we should be prepared to view it as a gradient rather than as a dichotomy between two homogeneous categories' (1985: 1257).

(3.48) She has three sons _ she could rely on for help, so she was not unduly worried.
(taken from Huddleston 1984: 399)

Even if the mother in question in (3.48) has exactly three sons so that the relative clause cannot be said to restrict the antecedent's reference, the relative clause cannot be omitted without severely altering the meaning of the entire sentence. The relative clause is clearly an 'integral' (i.e. essential) element of the DP: without it the remaining phrase *three sons* would lead to a radically different sentence meaning (cf. Schmied 1991a: 50f.). Thus instead of speaking of a restrictive or non-restrictive dichotomy, it seems more appropriate to distinguish obligatory and optional relative clauses.

[6] Comparing (3.47a) and (3.47b) also illustrates another difference between restrictive and non-restrictive clauses: being optional, the latter obviously have weaker semantic ties with the antecedent. This is indicated by the orthographical convention of putting the non-restrictive relative clauses in commas, which itself mirrors prosodic effects, since in speech non-restrictive relative clauses are often separated from their antecedent by a pause (Huddleston, Pullum and Peterson 2002: 1058).

Using optionality as his defining criterion,[7] Schmied (1991a) tested the influence of the relative junction on the choice of the relativizer, drawing on British English (Lancaster–Oslo–Bergen) and Indian English (Kolhapur) corpora. He found that *Ø* is in fact limited to obligatory contexts. Furthermore, *that* is also mostly used in obligatory relative clauses, although it occurs in a few optional ones. In contrast, the *wh*-pronouns are widely employed in all types of relative clauses (cf. Schmied 1991a: 197).

Thus, summarizing the above results, it becomes apparent that an analysis of the two relativizers obligatorily demanding stranding will have to take into account that *Ø* and *that* are strongly restricted to obligatory relative clauses. An interesting question arising from this observation is whether these secondary influences also have consequences for preposition placement in *wh*-relative clauses. If *that* and *Ø* are (more or less) limited to obligatory contexts, then it might be that in obligatory *wh*-relative clauses stranding will also be more frequent since there exists a recurrent overt 'relativizer+stranded preposition' model exerting an analogical influence. Alternatively, it is also conceivable that there are factors which independently favour both *that/Ø* and preposition-stranding. According to this line of reasoning, stranding might then be less frequent with *wh*-relativizers in obligatory relative clauses since a factor favouring stranding might also lead to a preference of *that/Ø* over a *wh*-relativizer. In optional relative clauses, on the other hand, *that* and *Ø* are not available choices. Hence in optional relative clauses *wh*-relativizers might occur more frequently in contexts favouring stranding, thus causing optional relative clauses at one point to be identified as favouring stranding with *wh*-relativizers. Drawing on the same input it is thus possible to derive opposite predictions concerning the influence of the factor restrictiveness on preposition placement in *wh*-relative clauses. This paradoxical situation requires that the relative clause corpus data alone must be subjected to an additional multivariate corpus analysis (since restrictiveness is a factor that only applies to relative clauses): it must be investigated whether the restrictiveness of the relative clause has an independent effect on preposition placement in *wh*-relative clauses.

The advantage of such a multivariate analysis is, of course, that it allows explicit testing of the independence of an effect. To illustrate this point, take Ungerer *et al.*'s claim that '[a]s non-defining relative clauses are mainly used in formal English, the preposition normally comes before the relative pronoun [in these formal non-defining RCs]' (1996: 206). According to Ungerer *et al.*, the restrictiveness of a clause has no independent influence on the preposition placement, and the increased percentage of pied-piped tokens in

[7] Schmied actually subdivided relative clauses into four obligatory and two optional categories. Emphasizing the impossibility of a clear-cut categorization of all relative clauses in terms of their relative junction, he also had an extra class for 'ambivalent' cases: 'The relationship between the dichotomy obligatory–optional [is] more a cline than absolute …' (1991a: 74).

non-restrictive relative clauses can simply be attributed to other factors, e.g. the level of formality. While this might be the case, only a statistical multivariate analysis of the corpus data allows assessment of the actual influence of apparently influential factors.

Apart from the restrictiveness of the relative clause, several studies, such as Ball's (1996) or Guy and Bayley's (1995), have tried to show that the choice of the relative pronoun is also influenced by various other factors such as antecedent type (personal vs non-personal) or level of formality: '*wh*- forms are favoured in formal writing and for human antecedents in embedded-clause position. Choice of *that* is favoured in informal speech and for non-human antecedents. Zero is moderately favoured for human antecedents, especially in embedded-clause direct-object position, and in informal speech …' (Guy and Bayley 1995: 155). All of these effects can be seen as potential secondary influences on preposition placement. Possible effects of the factor finiteness have already been hinted at above: in section 3.1.1, it was pointed out that non-finite relative clauses do not license *that*-relativizers (**the man that to talk to*) and require *wh*-relativizers to obligatorily pied-pipe prepositions (cf. *the man to whom to talk* vs **the man who*(*m*) *to talk to*). As I will argue below, diachronically the ban on *that* in non-finite relative clauses can be seen as a direct factor that resulted in the ban on stranding with *wh*-relativizers in these constructions. In any case, the relative clause corpus data was explicitly coded for the factor finiteness.

Concerning the type of antecedent it was expected that this factor has no secondary influence on preposition placement. It was therefore decided not to directly code the corpus data for the type of antecedent. Note that it was still possible to detect potential effects of this factor: if the type of antecedent should cause *who*(*m*) and *which* to behave differently with respect to preposition placement, these effects should be mirrored somehow by Ø and *that*, which are also favoured by human and non-human antecedents, respectively.

Finally, Guy and Bayley (1995) showed that *that* and Ø are favoured by informal contexts, while *wh*-pronouns are preferred in formal contexts. This is thus a case were it is not easy to distinguish primary from secondary effects: as was indicated above (and will be discussed in detail below), preposition-stranding is favoured in informal contexts. In addition to this, informal contexts also favour *that* and Ø, which obligatorily lead to stranding. The situation in relative clauses is therefore far more complex than in, for example, interrogative clauses: while the level of formality in interrogative clauses only affects preposition placement, in relative clauses it influences preposition placement and the choice of relativizer (which themselves are not independent). The interesting point here is obviously the behaviour of the *wh*-words: do *wh*-relativizers, which in contrast to their interrogative counterparts are already associated with more formal registers, also cooccur more frequently with the more formal preposition placement variant, i.e. pied-piping? If so, is this an epiphenomenon of the formality of *wh*-relativizers or is it part of the speaker's mental grammar?

All of the above facts led to the decision to subject the relative clause corpus data to an additional multivariate analysis which also included factors such as the restrictiveness and the finiteness of the relative clause as well as the level of formality (in fact all corpus data were coded for this latter factor; for details see below).

3.1.3 Different variable effects across clause types

In the preceding chapters we have already come across a case where clause types which are generally taken to allow both stranding and pied-piping differ with respect to preposition placement: while non-finite *wh*-interrogative clauses allow preposition-stranding (e.g. *I wonder who to talk to*) in English, in non-finite *wh*-relative clauses stranding is prohibited (cf. **the man who to talk to*). Besides such categorical effects, however, the question is whether individual clause types also exhibit idiosyncratic variable constraints on preposition placement. In *wh*-relative clauses numerous primary and secondary effects were seen to be operating (see section 3.1.1 and 3.1.2.2). It would therefore not be surprising to find that *wh*-relative clauses have different preposition placement preferences compared to, for example, *wh*-interrogative clauses, which are not subject to the same types of constraints.

First of all, it needs to be seen whether all clause types inducing obligatory preposition-stranding also license the same types of stranded prepositions:

(3.49) a. John hated **the jokes**$_i$ [Ø **(that)**$_i$ Bob laughed **at**$_i$]. [non-*wh*-relative]
b. John laughs at **the same jokes**$_i$ as [Ø I laugh **at**$_i$]. [comparative]
c. **His jokes**$_i$ were impossible [Ø to laugh **at**$_i$]. [hollow]
d. [**His jokes**$_i$ have been laughed **at**$_i$ enough.] [passive]
e. [**What**$_i$ he is laughing **at**$_i$] is called stand-up comedy. [free relative]

From a Minimalist perspective the constructions in (3.49) are actually a mixed bag: the sentences in (3.49a–c) are generally taken to involve movement of an empty operator *Ø* within the embedded clause (Grewendorf 2002: 76; Radford 2004: 233f.). Consequently, the co-indexed NPs headed by *joke* in these examples do not enter the derivation as the complement of the preposition *at* and therefore cannot pied-pipe it. In contrast to this, both co-indexed elements in (3.49d,e) are considered to start off as the complement of *at*, but then move to positions within different functional projections. In passive sentences the NP moves to a position that can normally only be filled with arguments, the specifier of the T(ense) projection (cf. Chomsky 2000: 102; Radford 2004: 241–80). In free relative clauses the *wh*-word moves into a position somewhere in the C(omplementizer)-system (expressing force/mood) (Chomsky 2000: 102; as Radford (2004: 233) points out, the final landing site of the *wh*-word is heavily disputed in Minimalist analysis of free relative clauses; for an overview see Bury 2003 and van Riemsdijk 2000).

While I do not necessarily share several of the underlying assumptions of the Minimalist Program, it is at least interesting to see that in this approach the structures in (3.49) are regarded as being the result of different operations. Consequently one might wonder whether these different operations are viewed as having different effects on the types of stranded prepositions which are licensed. As it turns out no particular idiosyncratic effect of non-*wh*-relative, comparative, hollow and free relative clauses seems to have been postulated. Passive sentences, however, are well known to place tighter restrictions on stranded prepositions than the other clause types: basically, only two types of prepositions are said to be stranded in prepositional passives such as (3.49d): (1) prepositions which are lexically specified by a verb (e.g. *approve*, which obligatorily selects *of* (3.50a)), or a verbal idiom (e.g. *lose sight of* in (3.50b), which is also a stored complex lexical item), and (2) locative prepositions which are not specified by a verb (cf. (3.51), in which *on* is not obligatorily associated with *sit*; see Quirk *et al.* 1985: 163; Ward, Birner and Huddleston 2002: 1433):

(3.50) a. **His plan** has been **approved of**.
b. **The great plan** has been **lost sight of**.

(3.51) **This chair** has been sat **on**.

What both types of prepositional passives have in common is, of course, the fact that the NP which logically functions as the prepositional complement is an affected participant of the verbal action (cf. Quirk *et al.* 1985: 1165). This accounts for the fact that a locative adjunct PP such as *on this chair* can appear in a prepositional passive as in (3.51), but temporal adjunct PPs such as *on Mondays* cannot (cf. **Mondays were worked on*; cf. Ward, Birner and Huddleston 2002: 1434): while a chair can be said to be affected by someone sitting on it (it might be warm for some time even after the person has left; there might be dents or wear and tear effects visible), days of the week are abstract concepts that are not affected by whatever people do. In addition to this, transitive prepositional verbs (i.e. verbs which obligatorily select for an object and a prepositional phrase) do not permit prepositional passive (cf. (3.52)):

(3.52) *__I__ was explained the problem **to**.
(Ward, Birner and Huddleston 2002: 1433)

In contrast to the different underlying syntactic operations in the derivation of the obligatorily stranding clause types, the clauses exhibiting variable preposition placement are usually captured by a single mechanism, i.e. *wh*-movement to the SpecC position (Radford 2004: 188–240). Consequently, preposing,[8] interrogative, exclamative and *wh*-relative clauses might be expected to behave alike with respect to preposition placement:

(3.53)	a.	**[His joke]**$_i$ I laughed **at**$_i$.	[preposing]
	b.	**[What]**$_i$ did he laugh **at**$_i$?	[interrogative]
	c.	**[What a dumb joke]**$_i$ he laughed **at**$_i$!	[exclamative]
	d.	the joke [**[which]**$_i$ he laughed **at**$_i$]	[*wh*-relative]
(3.54)	a.	**[At his joke]**$_i$ I laughed$_i$.	[preposing]
	b.	**[At what]**$_i$ did he laugh$_i$?	[interrogative]
	c.	**[At what a dumb joke]**$_i$ he laughed$_i$!	[exclamative]
	d.	the joke [**[at which]**$_i$ he laughed$_i$]	[*wh*-relative]

Yet, frequently a distinct difference between *wh*-relative and *wh*-interrogative clauses is postulated in the literature: while, for example, Van den Eynden (1996: 444) claims that 'stranding is not really an option with WH-… relatives; not now and not in the past', Visser (1963: 406) maintains that '[w]hen an interrogative sentence or dependent clause opens with *whom* or *what* the preposition has end-position. The putting the preposition before these pronouns has always been less usual; today this usage is reserved for literary diction'.

As Trotta points out, one reason why interrogative clauses might strongly favour preposition-stranding is their discourse function: in interrogative clauses 'the *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' (Trotta 2000: 55). Since *wh*-interrogative words should therefore introduce questions, pied-piping is dispreferred and the preposition is left stranded. Therefore (3.53b) should be preferred over (3.54b). In bound relative clauses, on the other hand, 'the *wh*-word does not represent unspecified information: the antecedent precedes the *wh*-XP and in effect "signposts" that something else is coming, which, since something else is known, may be delayed over a longer stretch of language' (Trotta 2000: 55). According to this view, pied-piping should be far more frequent in relative clauses such as (3.54d) since the antecedent noun (*joke* in (3.54d)) is known information which allows a preposition to intervene between it and the relativizer. Furthermore, another advantage of the pied-piped preposition in relative clauses is that it also adds information as to the syntactic and semantic function of the *wh*-word in the relative clause. Thus a pied-piped preposition might potentially facilitate the interpretation of the entire relative clause.

[8] Though see Grewendorf (2002: 75f.) for an analysis in which topicalization in English involves movement of an empty operator.

Table 3.1 *P placement in* wh-*interrogative and* wh-*relative clauses in the Brown corpus (Trotta 2000: 57, 182)*

Clause	stranded	pied-piped	overall
Interrogative	100	57	157
Relative clause	12	1054	1066
overall	112	1111	1223

Each cell has a χ^2 value larger than 6.63, i.e. a probability of $p < 0.01$. Cells whose observed frequency is lower than their expected frequency have been shaded light grey, while those with an observed frequency higher than expected have been shaded dark grey.

So far no synchronic study that I am aware of has statistically tested the hypothesis that *wh*-relative and *wh*-interrogative clauses exhibit different preposition placement preferences. Nevertheless, there have at least been some attempts at empirically investigating the matter. Trotta (2000), for example, examined the distribution of preposition placement in the Brown corpus, a one-million-word corpus consisting of written texts of American English from the 1960s (cf. McEnery and Wilson 1996: 185). Table 3.1 gives an overview of his results.

Trotta, unfortunately, does not subject his results to a rigorous statistical analysis, yet his findings are already indicative. As a simple chi-square (χ^2) test shows, the distribution in Table 3.1 is highly significant ($\chi^2 = 636.468$, df = 1, $p << 0.001$ for a Pearson's chi-squared test with Yates's continuity correction for 2×2 tables). In addition to this, the Cramer's ϕ value of 0.726 for this chi-square test indicates that the association of the factors CLAUSE TYPE and PREPOSITION PLACEMENT seems particularly strong (since a ϕ value of 0 indicates no correlation, while a value of 1 would be a perfect correlation; see Gries 2008: 178–9). The disadvantage of such a chi-square test is, of course, that it only tests the influence of one factor group. It is still possible that the figures in Table 3.1 are simply the result of a complex factor association. If, for example, all stranded interrogative tokens came from informal text types, while all pied-piped relative clause tokens were taken from formal text types, then the apparent effect of the factor clause type could in fact be due to an effect of the level of formality. While this again shows the need for a multivariate analysis of preposition placement, the data in Table 3.1 at least appear to corroborate the claim that *wh*-interrogatives favour stranding, while *wh*-relative clauses favour pied-piping.

If a distinct difference with respect to stranding and pied-piping can be assumed for *wh*-relative and *wh*-interrogative clauses, the question arises whether preposed/topicalized (3.53a/3.54a) and exclamative (3.53c/3.54c) clauses also show signs of preposition placement preferences. For exclamative sentences Huddleston claims that 'the exclamative feature percolates upwards in the same way as the interrogative feature' (2002b: 918). But why should exclamative and interrogative clauses behave alike with respect to pied-piping? Trotta offers the following explanation:

> the fronted Exclam[ative]P[hrase] has a special thematic significance in English *wh*-exclamative clauses; the 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. (Trotta 2000: 107)

Thus, there are functional reasons why the *wh*-exclamative markers *how* and *what* need to introduce exclamatives. Consequently, just as in interrogative sentences, pied-piping should be strongly dispreferred. This would explain Quirk *et al.*'s observation (1985: 834) that pied-piping with exclamatives is a rare phenomenon. In fact, Trotta goes as far as arguing that '[a]s regards pied-piping *vs* stranding, it seems reasonable to assume from the evidence that stranded prepositions are heavily favoured [in exclamatives] regardless of any grammatical, contextual or situational factor' (2000: 107).

From a functional perspective, preposing material to the front of the clause as in (3.53a) / (3.54a) can be said 'to draw the hearer's attention to an entity about which some new information is to be provided' (Jackendoff 2002: 412). While clause-type-specific restrictions on preposition placement are not mentioned in the literature, it is well known that discourse pragmatic constraints affect topicalization in general:

(3.55) a. **[To the girl]**$_i$ I talked$_i$.
b. **[The girl]**$_i$ I talked **to**$_i$.

In order for either the PP *to the girl* or the NP *the girl* to be preposed in (3.55) they must be discourse salient, i.e. the girl in question must have been mentioned before or must somehow be retrievable from the context (see e.g. Ward, Birner and Huddleston 2002: 1372). In addition to this, the difference between (3.55a) and (3.55b) can also be attributed to discourse-related factors: in (3.55a) the new information is that the speaker talked to the girl (but did not, e.g., kiss her), while in (3.55b) it is stressed that he talked to the girl but not necessarily to any other person present. What is important to note about this phenomenon is that unlike with *wh*-relative, *wh*-interrogative and exclamative clauses, in topicalized clauses there do not appear to be functional factors that favour either stranding or pied-piping.

This section has shown that clause-specific constraints appear to affect preposition placement in present-day English. As was emphasized, a considerable number of these constraints were attributed to functional factors. From a cognitive perspective the precise nature of these functional effects

Table 3.2 *P placement in wh-interrogative and wh-relative clauses in selected languages (adapted from Hamilton 1994)*

Language	Preposition-stranding	Pied-piping
Arabic	no	yes
Dutch	yes	yes
Finnish	no	yes[a]
French	no	yes
German	no	yes
Italian	no	yes
Korean	no	no
Mandarin	no	no
Russian	no	yes
Spanish	no	yes

[a] The claim that Finnish has pied-piping can only be entertained 'in the sense that case endings on the relative pronouns carry the preposition's meaning' (Hamilton 1994).

on mental grammars obviously needs to be assessed: do these functional factors universally affect all languages to the same degree? If so, the grammars of L1 and L2 speakers should be subject to the same constraints and the same effects should surface in British as well as Kenyan English. On the other hand, it is conceivable that these effects are universally present but that the degree to which they are grammaticalized differs from language to language. In order to answer this question it is necessary to see how preposition placement is constrained across clause types in different languages as well as how it developed in earlier stages of the English language.

3.1.4 Diachronic and typological evidence

Cross-linguistically preposition pied-piping is far more common than stranding (see e.g. Hawkins 1999: 277). A survey on the *Linguist List* (Hamilton 1994), for example, yielded the information summarized in Table 3.2. This shows that only Dutch is similar to English in licensing both preposition-stranding and pied-piping (but cf. fn. 9 below), while Korean and Mandarin do not allow either of the two options. The remaining languages only permit pied-piped prepositions. Since English historically is a West Germanic language, the case of German is particularly interesting. Take, for example, the following sets of sentences:

(3.56) a. *das Phänomen, [(**dass**)$_i$ er **über**$_i$ sprach] [non-*wh*-relative]
'the phenomenon **that** he spoke **about**'
b. ***das Gleiche**$_i$ wie [ich **über**$_i$ **gesprochen** habe] [comparative]
'**the same stuff** as I have **talked about**'

	c.	***Sein Buch**$_i$ war einfach [**über**$_i$ zu reden] ‘**His book** was easy to talk **about**’	[hollow]
	d.	***Fussball**$_i$ ist genug **über**$_i$ gesprochen worden ‘**Football** has been spoken **about** enough’	[passive]
(3.57)	a.	***[Ihm]**$_i$ habe ich schon **von**$_i$ gehört ‘**Him** I have heard **of**’	[preposing]
	b.	***[Wen]**$_i$ spricht er **über**$_i$? ‘**Who(m)** is he speaking **about**?’	[interrogative]
	c.	***[Was für eine Ausdauer]**$_i$ er **über**$_i$ verfügt! ‘**What stamina** he **has got**’ ([über Ausdauer]$_{PP}$ verfügen = ‘have stamina’)	[exclamative]
	d.	*das Phänomen, [**[das]**$_i$ er **über**$_i$ sprach] ‘the phenomenon **which** he spoke **about**’	[*wh*-relative]
(3.58)	a.	**[Von ihm]**$_i$ habe ich schon gehört$_i$. ‘**Of him** I have heard.’	[preposing]
	b.	**[Über wen]**$_i$ spricht$_i$ er? ‘**About who(m)** is he speaking?’	[interrogative]
	c.	**[Über was für eine Ausdauer]**$_i$ er verfügt$_i$! ‘**What stamina** he has got!’	[exclamative]
	d.	das Phänomen, [**[über das]**$_i$ er sprach$_i$] ‘the phenomenon **about which** he spoke’	[*wh*-relative]

German thus has none of the constructions which induce obligatory preposition-stranding in English (3.56a–d). In addition to this, even in preposing, interrogative, exclamative and *wh*-relative clauses pied-piping is obligatory in German[9] (cf. (3.57a–d) vs (3.58a–d)).

The German examples in (3.57b/3.58b) and (3.57c/3.58c) are particularly remarkable: remember that in the preceding chapter it was claimed that preposition-stranding is preferred in English *wh*-interrogative and exclamative clauses because of functional constraints: since the *wh*-word indicates the function of the clause in both cases, it should come as early as possible and a preceding pied-piped preposition is therefore dispreferred in these structures. As the German data show, however, the function of *wh*-interrogative and exclamative clauses is not a factor that necessarily leads to preposition-stranding. It only assumes relevance once the grammar of a particular language generally licenses stranded prepositions. This observation

[9] Note, however, that there are constructions in which pronominal adverbs exhibit a phenomenon similar to preposition-stranding:

(i) Preposition-stranding?
a. **Da** will ich nichts wissen **von**! (‘**That** I don’t want to know anything **about**’)
b. **Davon** will ich nichts wissen! (‘**That-about** I don’t want to know anything’)
c. **Da** will ich nichts wissen **davon**! (‘**That** I don’t want to know anything **that-about**’)

As (ib) shows, however, (ia) cannot be classified as preposition stranding. Instead, a postpositional morpheme is left without complement (*-von* in (ia)). Occasionally, the two constructions then compete with a third variant in which the initial pronominal morpheme appears at the front of the clause and the full pronominal adverb resurfaces later in the clause as in (ic). The phenomenon is therefore not treated on a par with preposition-stranding in English. Moreover, just as in German, stranding in Dutch (cf. Table 3.2) also seems to be limited to pronominal adverbs (see Fleischer 2002; Hoekstra 1995).

will obviously have to be kept in mind for the grammatical analysis of preposition placement in English.

Furthermore, it was argued in section 3.1.1 that *wh*-words in English free relative clauses cannot pied-pipe prepositions because they are a combination of head and relativizer. Despite the theoretical appeal of this explanation it needs to be pointed out that there seem to be occasional instances of pied-piping in German free relative clauses (cf. Kubota 2003; Müller 1999):

(3.59) a. [[**Aus wem**]$_i$ noch etwas herausgequetscht$_i$ werden kann], ist sozial dazu verpflichtet, es abzuliefern …
'Those who have not yet been bled dry are socially compelled to hand over their last drop' (Wiglaf Droste, *taz*, 01.08.97, p. 16; cited in Müller 1999: 61)
b. *[[**Wem**]$_i$ noch etwas **aus**$_i$ herausgequetscht werden kann], ist sozial dazu verpflichtet, es abzuliefern …

Example (3.59a) is an attested example from one of the larger German newspapers in which the pied-piped preposition *aus* introduces a free relative clause that acts as the subject of the entire sentence. As illustrated by (3.59b), German does not even allow preposition-stranding in cases which in English are deemed to obligatorily disallow pied-piping due to universal syntactic constraints. This is not to say that German generally licenses pied-piping in free relative clauses. In fact, the *wh*-word in German is normally required to show a morphological form that is compatible with the case requirements of both the verb of the free relative clause and the matrix verb (cf. *Er hilft*$_{\text{DATIVE}}$, *wem*$_{\text{DATIVE}}$ *er vertraut*$_{\text{DATIVE}}$ 'He helps whomever he trusts' vs **Er vertraut*$_{\text{DATIVE}}$, *wen*$_{\text{ACCUSATIVE}}$ *er kennt*$_{\text{ACCUSATIVE}}$ / **Er vertraut*$_{\text{DATIVE}}$, *wem*$_{\text{DATIVE}}$ *er kennt*$_{\text{ACCUSATIVE}}$ 'He trusts whomever he knows'; Müller 1999: 62). Nevertheless, data such as (3.59a) indicate that pied-piping in free relative clauses is not a structure that is typologically completely impossible. Again, this is a finding that will have to be incorporated into the grammatical analysis of preposition placement in English.

Yet why should pied-piping be typologically preferred over stranding as indicated by Table 3.2? Hawkins (1999, 2004) argues that this is due to the fact that from a processing perspective preposition-stranding is far more complex than pied-piping. First of all, preposition-stranding can give rise to garden path effects, while pied-piping avoids such on-line misanalyses (examples taken from Hawkins 1999: 277):

(3.60) a. [Which student]$_i$ did you ask (O$_i$) Mary about O$_i$
b. [About which student]$_i$ did you ask Mary O$_i$

After having processed the main verb *ask* the human processor is prone to assign *which student* as the filler of the object gap in (3.60a), which leads to a garden path effect once *Mary* is encountered. The pied-piped alternative in (3.60b), on the other hand, does not yield such an effect. Pied-piping thus complies with Hawkins's 'Avoid Competing Subcategorizors' principle which states that '[t]he human processor prefers to avoid garden paths that result from competing subcategorizors within [a Filler–Gap Domain]' (1999: 277).

Secondly, as (3.60b) shows, pied-piping also has the advantage that the filler only has to be identified with a gap within the VP, while in (3.60a) the gap is embedded within a PP that itself is embedded in the VP. Thus in (3.60b) the filler can be successfully integrated upon processing the main subcategorizor of the clause (i.e. the main verb *ask*), while in (42b) this integration is deferred. This again conforms to one of the processing principles postulated by Hawkins, namely 'Valency Completeness': 'The human processor prefers [Filler–Gap Domains] to include the subcategorizors for all phrases within the domain that contain the gap' (Hawkins 1999: 278; cf. also 2004: 210–5).

According to Hawkins, processing preferences can therefore explain why pied-piping should generally be preferred over stranding. However, as he also notes, prepositions which 'are highly dependent on verbs for their interpretation and processing' (Hawkins 1999: 260, fn. 15) can be processed far more easily if remaining *in situ*. Furthermore, he assumes that 'the ratio of stranding to pied-piping in English should be proportional to the degree of dependency between V and P' (Hawkins 1999: 260, fn. 15). Consequently, lexemes such as *rely on* in which the preposition *on* obligatorily co-occurs with *rely* should easily be stranded: in structures like *the man who he relied on* the preposition is already expected by the processor after encountering *relied*, so that preposition-stranding with such prepositional verbs does not result in a great increase in processing cost. While this effect in present-day English will be examined in section 3.2 ('Type of PP'), data from Old English (OE) already corroborate the claim that stronger verb–preposition dependencies favour stranding.

In OE a 'stranded preposition is always part of a complement PP, and it nearly always immediately precedes the verb. In those cases where it is not immediately left, it is right adjacent to the verb' (Fischer *et al.* 2000: 158). In addition to these constraints, which indicate that the closeness of the main subcategorizor and a preposition was already of vital importance, the following examples show that OE already exhibited many of the categorically stranding contexts found in present-day English:

(3.61) [non-*wh*-relative]
On ðam munte Synay, **ðe** se Ælmihtiga **on** becom […]
on the mountain Sinai **that** the Almighty **on** came […]
'On mount Sinai **on which** the Almighty came'
(Ælfric's Catholic Homilies II, 12.1.116.226; cited in Fischer *et al.* 2000: 60)

(3.62) [comparative]
[…] to **beteran tidum** Þonne we nu **on** sint.
[…] for **better times** than we now **in** are
'for **better times** than we are now **in**'
(Ov 25.86.2; cited in Traugott 1992: 264)

(3.63) [hollow]
Seo burg […] wæs swiðe fæger **an** to locianne
the castle […] was very fair **at** to look
'**The castle** […] was very fair to look **at**'
(Ælfred, Oros. (Sweet) 74, 13; cited in Visser 1963: 397)

(3.64) [passive]: **not attested in OE**

(3.65) [free relative]
And heo gefret softnysse oððe sarnysse, **swa hwæð swa** heo **on** bið.
and it feels softness or pain **so which as** it **in** is
'And it feels softness or pain, **whichever** it is **in**'
(Alc.P.XI.218; cited in Allen 1980: 279)

The first point to note is that OE does not seem to have had a prepositional passive. In fact, the construction does not surface in texts until the thirteenth century, i.e. the Middle English (ME) period. Denison (1985, 1993) argued that the pattern spread via lexical diffusion starting first with 'combinations of verb and prepositions that semantically form a close unit, such as *lie by* "sleep with" and *set of/let of/tell of* "regard"' (Fischer *et al.* 2000: 78). The non-compositionality of these verb– preposition lexemes thus seems to have opened up the possibility of using the NP following the preposition as the subject of a passive sentence (cf. Denison 1993: 141; Fischer *et al.* 2000: 78). An underlying V [P NP] order was thus 'reanalysed' as a [V P] [NP] sequence (e.g. Fischer 1992: 386). In the ME example given in (3.66), for example, the NP *sche* thus could become the subject of a passive clause since *lie by* was interpreted as a single lexeme:

(3.66) [passive]
[…] **sche** was yleyen **bi**
[…] **she** was lain **with**
'**she** had been lain **with**'
(Arth. and M. (Auch) 849; cited in Denison 1993: 125)

Once the number of idiomatic prepositional verbs started to increase significantly in late ME and Early Modern English (EModE), so did instances of preposition-stranding in passives (Denison 1993: 143). The spread of preposition-stranding was then further supported by the emergence of phrasal verbs like *switch on* or *give up*. As (3.67) shows, phrasal verbs are complex verbs which consist of a verb (e.g. *switch*) and a particle (e.g. *on*). Even though transitive phrasal verbs appear to be similar to V-PP combinations (3.67a), a main difference concerns the fact that particles, unlike prepositions, can also follow an object NP ((3.67b); cf. Quirk *et al.* 1985: 1156ff.):

(3.67) a. She **switched on** the light
b. She **switched** the light **on** (Quirk *et al.* 1985: 1157; emphasis added)

Example (3.67b) thus illustrates that the particle and the object NP do not form a constituent. Consequently, the underlying structure of (3.67a) would be something like [V PRT] [NP], i.e. $[\textit{switch on}]_V$ $[\textit{the light}]_{NP}$. This analysis explains why the particle of a phrasal verb can never pied-pipe to an object (cf. Bailey 1986: 156). Since the *wh*-object *which* and the particle *away* are not a constituent, they cannot be moved together to the front of a clause as in (3.68b). As a result, particles are always 'stranded':

(3.68) a. I washed the dishes **which** John had **put away**
b. *I washed the dishes **away which** John had **put**

The underlying $[\text{V PRT}]_V$ structure of examples like (3.68a) is then said to have reinforced the reanalysed [V P] complexes arising in ME. Note, however, that while the rise of prepositional verbs and phrasal verbs can be said to have helped the spread of the prepositional passive, this alone does not explain the diachronic development of preposition-stranding in English. As can be seen in the sentences in (3.61–3.65), even before the rise of prepositional verbs and particle verbs, OE had obligatory stranding in non-*wh*-relative (3.65), comparative (3.62), hollow (3.63) and free relative clauses (3.64). Selectively, OE grammar had therefore already licensed stranded prepositions which were adjacent to the main subcategorizor of the clause.

Turning to clause types with variable preposition placement in modern-day English, it is often claimed that in OE preposed or 'topicalisation constructions preposition stranding is not found except when the topicalised NP is a personal or locative (*ðær*, *ðyder*) pronoun' (Fischer 1992: 389). Yet, while the majority of topicalised tokens are pied-piped (as in (3.69)), Denison (1985: 192) also gives an attested OE stranded example (3.70):

(3.69) [topicalization pied-piped]
To ðæm soðum gesælðum ic tiohige ðæt ic ðe læde
'**To the true happiness** I intend (that) I thee lead'
(Boeth.XXII.2 p.51.12; cited in Allen 1980: 286)

(3.70) [topicalization stranded]
Freond ic gemete **wið**
friend I meet **with**
'May I meet with a friend'
(*MCharm* 11.37; cited in Denison 1985: 192)

The literature contains no information on preposition placement in OE exclamatives (Allen 1980; Denison 1985, 1993; Fischer 1992; Fischer *et al.*

2000; Visser 1963), but in *wh*-interrogative (3.71) and *wh*-relative (3.72) clauses pied-piping was obligatory (Fischer *et al.* 2000: 55, 59, 92):

(3.71) [interrogative]
To hwæm locige ic buton to ðæm eaðmodum?
'**To whom** shall I look but to the humble?'
(CP Sweet p. 299.19; cited in Allen 1980: 285)

(3.72) [*wh*-relative]
Þæt fyr getacnode ðone Halgan Gast, **ðurh ðone** we beoð gealgode
'The fire betokened the Holy Ghost, **through whom** we are hallowed'
(Ælfric's Catholic Homilies II, 17.167.190; cited in Fischer *et al.* 2000: 59)

Concerning the obligatory status of preposition pied-piping with *wh*- relative and *wh*-interrogative pronouns, the OE system is thus comparable to the German one.[10] Nevertheless, there is one important difference between the German and the OE relativizer system: whereas German only has relative pronouns (cf. (3.58d) vs (3.56a)), OE (cf. (3.61)) – just like present-day English – also employed indeclinable complementizers as relativizers, which exhibited the expected obligatory stranding of prepositions (see section 3.1.1):

(3.73) On ðam munte Synay, **ðe** se Ælmihtiga **on** becom …
on the mountain Sinai **that** the Almighty **on** came
'On mount Sinai, on which the Almighty came'
(Ælfric's Catholic Homilies II, 12.1.116.226; cited in Fischer *et al.* 2000: 60)

Dekeyser (1990: 94, 104) points out that English is the only modern Germanic language which uses both strategies, relative pronouns and indeclinable complementizers. The other Germanic languages either exclusively use relative pronouns (e.g. Dutch, German; cf. above and fn. 9) or complementizers (the Scandinavian languages, e.g. Swedish; note the expected obligatory stranding of the preposition in (3.74)):

(3.74) Swedish: complementizer + obligatory stranding
a. huset **som** jag bor **i**
'the house **that** I live **in**'
b. *huset **i som** jag bor
'the house **in that** I live'
(adopted from Dekeyser 1990: 103)

As a result of employing both types of relativizers, OE thus had both pied-piped prepositions (with relative pronouns) and stranded ones (with complementizers).

[10] Note that whereas OE, just like German, had 'D'- relative pronouns (cf. *ðone* in (3.72) and *das* in (3.58d), respectively), ME had a new set of relative pronouns, the *wh*-forms, which in OE were used only as interrogative pronouns (cf. *hwæm* in (3.71), which corresponds to present-day English *whom*). For a detailed history of the *wh*-forms see Fischer *et al.* (2000: 92ff.).

Now, from the thirteenth century preposition-stranding started to extend to *wh*-relative (3.75) and *wh*-interrogative clauses (3.76) and became more common towards the end of the ME period (Dekeyser 1990: 91–2; Fischer 1992: 390):

(3.75) for nadde they but a sheete, **Which** that they myghte wrappe hem **inne**
'for they had nothing but a sheet **which** they could wrap themselves **in**'
(Chaucer, *The Canon's Yeoman's Tale c.* 1395; cited in Bergh and Seppänen 2000: 298)

(3.76) 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)

As Fischer notes, the analogical stranding of prepositions first occurred in *which*-clauses. According to her, one reason for the extension of preposition-stranding to *which*-clauses was the fact that '*which* was virtually indeclinable' (1992: 391). Thus, lack of overt case-marking facilitated the identification of *which* with the indeclinable complementizer *that*. Once preposition-stranding had spread from*that*-relative clauses to *wh*-relative clauses such as (3.75), it also could affect *whom*-interrogative clauses (3.76).[11]

One advantage of such an analogy-hypothesis is that it can account for the obligatory preposition pied-piping in non-finite *wh*-relative clauses: since the finite complementizer *that* is banned from non-finite relative clauses, there never existed an 'overt relativizer+stranded preposition' model (i.e. **the man that to talk to*) in non-finite relative clauses. Therefore, there was no analogical model which could have affected the non-finite *wh*-relative clauses. Consequently, preposition piping remained obligatory there (*the man to whom to talk* vs **the man who*(*m*) *to talk to*).[12]

Nevertheless, even though the lack of overt case-marking seems to have favoured the analogical extension of stranding to *which*-clauses, it must be emphasized that even during these earlier stages of English Radford's (1997) hypothesis that case-marking obligatorily leads to pied-piping (see section 3.1.2.1) cannot be entertained. In his study on the distribution of *whom* and *who* in Early Modern English, Schneider (1992) found twenty-four instances of *who*(*m*) occurring with a stranded preposition in a corpus of Shakespeare plays and poems. As it turned out, 50 per cent of these cases exhibited the case marked *whom* (1992: 442).[13]

11 One problem with this hypothesis is that both *wh*-relative and *wh*-interrogative clauses exhibit preposition-stranding from the thirteenth century onwards (see e.g. Fischer 1992: 390). The diachronic textual evidence therefore does not support the successive spreading of stranding from relative to interrogative clauses.

12 Note that non-finite *wh*-relative clauses such as *the man to whom to talk* did not exist in OE, and only begin to surface at the end of the ME period (cf. Kjellmer 1988: 564). They thus not only lacked a '*that* + stranded preposition' model, but also were not part of the language when prepositions-stranding started to extend to *wh*-clauses.

13 It must be pointed out, however, that Schneider's corpus included relative and interrogative uses of *who*(*m*). Taking into account a variety of possibly influential factors, a closer statistical analysis of the data showed that stranded prepositions did in fact strongly

Table 3.3 *P placement in* wh-*interrogative and* wh-*relative clauses in selected clause types (adapted from Yáñez-Bouza 2004)*

Clause type	Preposition-stranding	Pied-piping	Total
Preposed	17	11	28
Interrogative	6	7	13
Exclamative	1	0	1
Wh-relative	54	477	531
Total	78	495	573

Once preposition-stranding had extended to clauses which exhibited categorical pied-piping in OE, the expected functional constraints discussed in section 3.2.1 finally seemed to take effect. An investigation of a 242,270-word corpus of EModE texts (*c*.162,270 words from the Helsinki Corpus of English Texts, Diachronic and Dialectal, 1500–1700 and *c*.80,000 words from the Century of Prose Corpus, 1700–1780) by Yáñez-Bouza (2004), for example, yielded the results in Table 3.3. Despite the fact that the usual proviso applies (i.e. that the data needs to be subjected to a multivariate statistical analysis), Table 3.3 appears to support the predictions: while pied-piping is strongly favoured in *wh*-relative clauses (with a frequency of 89.8%), both pied-piping and stranding appear more or less equally acceptable in *wh*-interrogatives (53.8% vs 46.2%, respectively). In preposed/topicalized structures, preposition-stranding is even preferred over pied-piping (with 60.7% vs 39.3%). Finally, the only attested example from an exclamative has a stranded preposition. The fact that several cells have expected frequencies below 5 precludes the possibility of subjecting the data in Table 3.3 to a chi-square test. Resorting therefore to a Fisher's exact test for count data shows that the clause-type effects exhibited by these data are indeed strongly significant ($p << 0.001$).[14]

Section 3.1 has shown how different clause types exert different effects on preposition placement: on the one hand, there are a number of categorical contexts (non-*wh*-relative, comparative, hollow, passive and free relative clauses). In the empirical analysis of these clause types the most pertinent question will obviously be whether they behave alike with respect to the stranded prepositions they license or whether they exhibit any idiosyncratic effects. On the other hand, it was seen that there are functional reasons which might explain preferences for pied-piping or stranding with those

favour *who* over *whom* (with a VARBRUL factor weight of .73 (Schneider 1992: 442). Nevertheless, the examples of an inflected *whom* co-occurring with a stranded preposition challenge the claim that case-checking automatically entails obligatory *Wh*-feature percolation.

14 I am grateful to the anonyomous reviewer who suggested running a Fisher exact test over these data.

clause types that basically allow for both options (i.e. that *wh*-interrogatives should favour stranding far more than *wh*-relatives). Furthermore, since relative clauses are subject to complex primary and secondary effects on preposition placement, it was argued that this clause type should receive particular attention. Finally, at least in the case of *who* and *whom* it was seen that the type of displaced element also needs to be taken into account.

In the following sections I will now turn to factors which have been claimed to further restrict stranding and pied-piping in those clauses which allow variable preposition placement. I will first turn to the closeness of the verb–preposition relationship since this has already been touched upon above.

3.2 Type of PP

3.2.1 PP complements vs PP adjuncts

In the literature on preposition placement, researchers usually emphasize a distinction between the behaviour of PP complements and PP adjuncts. Hornstein and Weinberg (1981), for example, contend that stranding is only possible with PP complements, while PP adjuncts are supposed to cause categorical pied-piping.

(3.77) John decided on the boat
(3.78) a. 'John decided while standing on the boat'
b. 'John decided to buy or look at the boat'
(3.79) The boat which John decided on

As (3.78) shows, the PP *on the boat* in sentence (3.77) is ambiguous: it can either be interpreted as a PP adjunct referring to the location of the event (3.78a) or as a PP object which is affected by the decision process (3.78b). According to Hornstein and Weinberg, however, the relative clause in (3.79) is unambiguous, the only grammatical reading available being the one in which the PP headed by *on* functions as an object, i.e. in which *the boat* is seen as an object affected by the decision process (Hornstein and Weinberg 1981: 58–9).

Apart from the fact that it is debatable whether a PP adjunct reading is really unavailable for (3.79), Hornstein and Weinberg's complement–adjunct distinction is also faced with several other problematic counterexamples. Trotta, for example, points out that there is a 'mixed' group of PPs expressing certain, specifiable semantic roles which are usually classified as adjuncts, but do nevertheless allow both preposition-stranding and pied-piping (3.80–3.83): 'These are PPs which have an affected preposition complement, locative PPs expressing goal or target, PPs of accompaniment, and instrumental PPs' (Trotta 2000: 182):

(3.80) locative PP (with affected preposition complement)
a. The grass ***on which*** they walked was just planted
b. The grass ***which*** they walked ***on*** was just planted

(3.81) locative adverbial PP (goal/target adjunct)
  a. The bank ***to which*** he was rushing was closed
  b. The bank ***which*** he was rushing ***to*** was closed
(3.82) accompaniment adverbial PP
  a. The group ***with which*** I came here has disappeared
  b. The group ***which*** I came here ***with*** has disappeared
(3.83) instrumental adverbial PP
  a. There are many tools ***with which*** you can skin a cat
  b. There are many tools ***which*** you can skin a cat ***with***
  ((3.80–3.83) taken from Trotta 2000: 183)

Furthermore, even temporal PP adjuncts seem to strand prepositions occasionally, as in *the day which he arrived on* (Quirk *et al.* 1985: 1254). This leads Trotta to claim that the complement–adjunct constraint on preposition placement is not a strictly categorical one, but should rather be conceived of as a continuum (2000: 59f.). His position is supported by Johansson and Geisler's study (1998; Table 3.4) on preposition-stranding vs pied-piping in *which*-relative clauses in several spoken corpora (the London-Lund Corpus (LLC), and the spoken component of the Birmingham Corpus (BIRM) and the British National Corpus (BNC); note that Johansson and Geisler give row percentages in italics and column percentages in roman font). As Table 3.4 shows, using a simple binary complement–adjunct distinction, Johansson and Geisler's results show that pied-piping is clearly favoured with adjunct PPs in all three corpora (the pied-piped:stranded ratios being: 161:3 LLC; 660:10 BIRM; 120:9 BNC), but definitely not obligatory. Furthermore, in complement PPs the prepositions are in fact extremely prone to stranding (with the pied-piped:stranded ratios: 29:47 LLC; 139:119 BIRM; 27:56 BNC), but again the result is not a categorical one.

3.2.2 Types of PPs: The complement–adjunct cline

Since Trotta claims that the intermediate 'mixed' PP group allows equally for both preposition-stranding and pied-piping, and since the complement/adjunct distinction 'itself is a scalar rather than simply binary' (Trotta 2000: 59), it was decided to devise a more fine-grained classification of PPs for the present study.

The first criterion used to subclassify the syntactic functions of PPs concerned their obligatoriness: while adjuncts are generally considered optional, the obligatoriness of a phrase identifies it as a complement (Huddleston 2002a: 221–2). As the following examples show, adjuncts are always optional (3.86), while complements can be either obligatory (3.84) or optional (3.85).

Table 3.4 *Preposition pied-piping vs stranding in speech (LLC, BIRM, BNC) (taken from Johansson and Geisler 1998: 70)*

LLC	PiedP	%	Stranded	%	Total
Prepositional object	29	15%	47	94%	76
Adverbial	161	85%	3	6%	164
Total	190	*79%*	50	*21%*	240
BIRM	**PiedP**	**%**	**Stranded**	**%**	**Total**
Prepositional object	139	17%	119	92%	258
Adverbial	660	83%	10	8%	670
Total	799	*86%*	129	*14%*	928
BNC	**PiedP**	**%**	**Stranded**	**%**	**Total**
Prepositional object	27	18%	56	86%	83
Adverbial	120	82%	9	14%	129
Total	147	*69%*	65	*31%*	212

(3.84)	a. She perused the report	b.	*She perused	[obligatory complement]
(3.85)	a. She read the report	b.	She read	[optional complement]
(3.86)	a. She left because she was ill (taken from Huddleston 2002a: 221)	b.	She left	[optional adjunct]

Therefore, even though obligatoriness cannot be considered a sufficient criterion to distinguish between complements and adjuncts, it is at least a necessary condition: if a phrase is obligatory, then it must be a complement (Huddleston 2002a: 221). Consequently, due to their obligatoriness, the following PPs must be complements (Hoffmann 2003, 2005):[15]

(3.87)	a. He relied on his mother	vs	*He relied
	b. Doughnuts consist of sugar	vs	*Doughnuts consist
	c. He slept with Sarah ('had sex with')	vs	*He slept
(3.88)	a. He lived in the sixteenth century	vs	*He lived
	b. I was in Rome	vs	*I was
(3.89)	a. I gave the book to John	vs	*I gave the book
	b. He put the book on the table	vs	*He put the book

Even though all the PPs in (3.87–3.89) are obligatory, they obviously differ with regard to their subcategorization status: whereas copular verbs like *live* and *be* simply require some sort of complement (cf. **He lived* vs *He lived*

[15] Note that contextual information sometimes allows ellipsis of contextually retrievable arguments, so that the answer to the question *How much money did you give to charity?* can be *I gave five dollars* (cf. McKercher 1996: 97f.). Nevertheless, as (3.89) shows, without context the goal PP is obligatory. Moreover, there does not seem to be a context that would license sentences like **He relied*.

happily ever after/alone/with his girlfriend/on the moon/in the sixteenth century), *rely* and *consist* only license PPs with *on* and *of*, respectively (**It consists with sugar and water*; see Huddleston 2002a: 220). Furthermore, *he slept* would be perfectly grammatical if simply having its basic stative meaning (i.e. 'sleeping'). However, if it is supposed to refer to sexual intercourse, as in (3.87c), *sleep* obligatorily needs to be constructed with a *with*-PP. As (3.87c) shows, in non-compositional V-PP combinations such as *sleep with somebody* the PP can be considered an obligatory complement, since its omission would lead to 'an unsystematic change in … meaning' (Huddleston 2002a: 220).

Due to their mandatory subcategorization relationship, V-P combinations such as *rely on* or *sleep with* 'have sex with' are often considered complex lexical items, so-called 'prepositional verbs'. Yet, as Quirk *et al.* point out (1985: 1163), prepositional verbs also exhibit several syntactic characteristics which seem to indicate that they are not complex lexical V heads, like phrasal verbs (e.g. *bring up*), but V-PP constructions:

(3.90) **On whom** did he **rely**?
(3.91) Rome also **relied more and more on provincials** to fill the ranks of the army. <ICE-GB:W2A-001 #87:1>
(3.92) a. A: Who did he rely on? B: (**On**) his mother
b. Did he rely on Bill or (**on**) Jeff?
c. He relies on Jeff more often than (**on**) Bill

Examples (3.90–3.92) show that the preposition *on* clearly forms a prepositional phrase with its complement, since it can be pied-piped to a *wh*-complement in questions (3.90), and allows the insertion of an adverbial between it and the verb (3.91). Moreover, the entire PP can be separated from the verb in several other constructions (cf. its optionality in 'responses [3.92a], in coordinate constructions [3.92b], or in comparative constructions [3.92c]'; Quirk *et al.* 1985: 1163). Since complex verbs like the phrasal verb *bring up* do not allow such structures (cf. e.g. **Up whom did she bring* or **Did she bring up Bill or up John?*), it can be argued that prepositional verbs like *rely on* are in fact verbs that obligatorily subcategorize for a complement PP headed by a specific preposition.

As the examples in (3.89) illustrate, the class of ditransitive verbs licensing complement PPs can also be divided into those which exclusively subcategorize for a particular preposition (the goal PP of the verb *give* is always headed by *to*; cf. (3.89a)), and those which only require the PP to qualify as a goal (cf. (3.89b) and *he put the book under the bed/in the oven*).

Yet, as pointed out earlier, obligatoriness cannot be considered a sufficient criterion for the identification of complements (cf. (3.85) *he read* [*the report*]). Therefore, it becomes necessary to take a closer look at the heterogeneous class of optional PPs:

(3.93) a. John talked **to Bill** about Bob
b. They worked **at the job**

(3.94) a. John died/sneezed/wept/exploded/apologized/laughed **in Rome** (adapted from Radford 1988: 235)
b. Bill killed the cat/thought about suicide/was drunk **on Saturday**

(3.95) a. He slept **in a bed**
b. He ran **to the church**
c. He committed the crime **with John**
d. He killed the cat **with a knife**

(3.96) The murderer kills his victims **[quickly/in a cruel way]**

Even though the omission of the PPs in (3.93) does not result in an unsystematic change in the verb meaning (cf. *he worked/talked*), the main verbs can nevertheless be said to subcategorize for the given prepositions: in *work at the job* (3.93b), for example, the PP introduced by *at* does not specify the location of the action, but rather provides a theme-like object. Verbs like *talk*, on the other hand, usually realize their theme as an *about*-PP (cf. (3.93a); also *he spoke about/thought about/dreamt about something*). Consequently, verbs licensing an optional PP theme argument must specify which theme-preposition they can co-occur with (cf. the theme PPs **he worked about the job* vs **he talked at Bob*). Furthermore, (3.93a) does not mean that John talked **at** Bill, but **with** him: the PP *to John* does not simply denote the goal of the event, but refers to John as also taking an active role in the communication. Therefore, the verbs in (3.93) play an important role for the identification of the semantic roles of the prepositional complements: '[t]he choice of the preposition is not arbitrary, but nor is its content sufficient to identify the role by itself' (Huddleston 2002a: 228). As a result, the PPs in (3.93) are often classified as complements (cf. Hornstein and Weinberg's analysis of *he talked to Harry about Bill*).

As (3.94) shows, the most prototypical adjunct PPs are optional locative and temporal expressions which can appear with virtually any verb, while always maintaining a constant, independent meaning (cf. Radford 1988: 235). These PPs do not add thematic arguments to a predicate, but instead 'localize' events on a time–location scale (Ernst 2002: 328; in Quirk *et al.*'s (1985: 511f.) terminology being 'sentence adjuncts'). One way to illustrate the relative independence of such PPs is the *This happened*-test (see Brown and Miller 1991: 90): assuming that events can be paraphrased by *this happened*, any PP in the original sentence that can combine with the paraphrase will have to be analysed as modifying the entire event:

(3.97) a. John talked. *This happened to John/about Bill
b. They worked. *This happened at the job

(3.98) a. John died/sneezed/wept/exploded/apologized/laughed.
This happened in Rome.
b. Bill killed the cat/left his family/thought about suicide/was drunk.
This happened on Saturday.

(3.99) The murderer killed his victims. This happened quickly/in a cruel way.

According to the results in (3.97–3.99), not only the locative and temporal PPs in (3.98), but also the manner adverbials in (3.99) modify entire events. This agrees with Ernst's (2002: 59) claim that manner adverbials compare events 'to other possible events of V-ing'. Therefore, the adverbials in (3.96) indicate that in comparison to other killings, the murders referred to are carried out comparatively fast or brutally. Furthermore, just as with locative and temporal adjuncts, the meaning of manner PPs in (3.96) appears to be independent from the event's predicate. In other words, the semantic role of the prepositional complement is exclusively determined by the preposition.

Interestingly, with the 'mixed' adjunct group introduced in the preceding section (cf. (3.95)), the theta role of the prepositional complement also seems to depend solely on the preposition: even without context, the (prototypical)[16] meaning of these PPs can be established: *in the bed* and *to the church* are locational PPs, *with John* expresses accompaniment and *with a knife* refers to an instrument. Nevertheless, the *this happened*-test indicates that the 'mixed' PPs do not modify events:

(3.100) a. He slept. *This happened in a bed
b. He ran. *This happened to the church
c. He committed the crime. *This happened with John
d. He killed the cat. *This happened with a knife

The mixed status of the PPs in (3.100) derives from the fact that, similarly to the adjunct PPs in (3.94), the semantic role of the prepositional complement is determined by the preposition, but the entire PP also establishes a thematic relationship with the predicate, just like the complement PPs in (3.93). *In a bed* (3.95a) does not merely specify a location, but also an object affected by the 'sleeping' event. Therefore, the 'mixed' adjuncts in (3.95) (Quirk *et al.* 'predication adjuncts'; 1985: 511f.) can be said to 'add participants to an event beyond the arguments of the predicate (which are also participants in the event)' (Ernst 2002: 131).

Concerning their syntactic behaviour, the 'mixed' PP group presents the greatest challenge to any binary complement–adjunct classification. First of all, in contrast to the other 'mixed' PP types (cf. (3.102)), locative PPs containing a prepositional complement affected by the event (e.g. *the bed* in (3.95a)) can become the subject of a passive sentence (cf. (3.101)), a feature usually associated with complements (Radford 1988: 233; Trotta 2000: 184):

[16] As always with semantic roles, the notion of 'prototypes' has to be stressed. It is, of course, conceivable that there is a context in which *with John* is used as an instrument (e.g. if in a fairy tale, a giant throws a person at another protagonist and kills him; see Fillmore 1977: 66). This, however, is definitely not the prototypical reading of *with John*.

(3.101) a. The bed was slept in
b. One of the benches had been sat upon <ICE-GB:W2F-005 #97:1>
(3.102) a. *The church was run to
b. *John was committed (the crime) with
c. *A knife was killed with

Trotta (2000: 184), however, claims that the *do-so* pro-form test indicates that affected locative PPs are adjuncts: since *do so* obligatorily replaces a verb and all its internal complements, it follows that combining this pro-form with an overt complement will produce an ungrammatical result, since all complements are already included in the pro-form (cf. Huddleston 2002: 223).

(3.103) a. *Jill keeps her car in the garage but Pam does so in the road
b. Jill washes her car in the garage but Pam does so in the road
(taken from Huddleston 2002: 223)

Since (3.103a) is ungrammatical, it follows that *in the garage* is a complement of *keep* already covertly included in the pro-form. In contrast to this, since the locative PP can combine with *do so* in the second example, it must be an adjunct in (3.103b), (this analysis is supported by the fact that the paraphrase *This happened in the garage/in the road* is only available for (3.103b)).

(3.104) a. ?Jill sleeps in the bed and Pam does so on the couch
b. ?Jill walked on the grass but/and Pam did so on the pavement
(3.105) a. *Jill ran to the church but Pam did so to the house
b. Jill killed a cat with Jack and Pam did so with Bob
c. He killed the bats with a knife and Pam did so with a gun

As the examples in (3.105) show, the *do-so* test also gives conflicting results for the PPs of the mixed group: the ungrammaticality of (3.105a) indicates that goal/source PPs qualify as more complement-like according to this test, while accompaniment (3.105b) and instrument (3.105c) PPs behave like adjuncts. Now, Trotta claims that affected locative PPs can be combined with a *do-so* pro-form, and thus, with regard to this particular test, should be classified as adjuncts (2000: 184). *Contra* Trotta, however, it must be argued that the sentences in (3.104) do not seem to be fully grammatical, which again would support an analysis of affected locative PPs as more complement-like.

Interestingly, as predicted by Trotta (2000: 184), the prepositional complement of goal/source PPs cannot be passivized on (see (3.102a)), but with regard to the *do-so* test, these PPs seem to behave like complements (cf. (3.105a)). The other two 'mixed' PPs, i.e. those expressing accompaniment or referring to instruments, on the other hand, are identified as adjuncts by both tests (cf. (3.102b,c); (3.105b,c)).

As the above results show, the class of 'mixed' PPs is an extremely heterogeneous one:

- due to the *do-so* and passivization tests, it seems possible to classify affected locations as complements. In addition, these PPs seem to be 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.
- Goal/source PPs are identified as complements by the *do-so* test. Furthermore, the only predicates licensing these PPs seem to be 'motion verbs' (e.g. *travel*, *go*, *fly*, *walk*, *run*, *roll*; cf. Larson 1988). Yet, these verbs do not appear to impose subcategorization restrictions on the prepositions (cf. *he ran into/towards/to/for the house*), which supports an adjunct analysis. Moreover, the theta role of the prepositional complement can be said to be exclusively assigned by the chosen preposition.
- Accompaniment and instrument PPs, on the other hand, fail all complement tests. With both types of PPs, the semantic role of the prepositional complement is mainly determined by the preposition. Yet, whereas the former can co-occur with a great range of verbs (e.g. *he laughed/ate/arrived/stood on the hill/killed the cat with his children*), the latter PP group appears to be restricted to actions implying or at least allowing for instruments (*he ate the apple/killed the cat/circumcised the baby with a knife* vs **he laughed with a knife*).

For the empirical analysis it will be interesting to see whether the fine-grained classification of PPs just outlined also has repercussions for preposition placement (i.e. which PP types show the same effects with respect to stranding and pied-piping). Besides the effect of the type of PP, however, the literature also contains claims as to the idiosyncratic effects of individual prepositions.

3.2.3 Idiosyncratic lexical effects

This section will discuss several idiosyncratic lexical effects on preposition-stranding and pied-piping which have been mentioned in the literature. Basically, three types of factors can be identified: (1) prepositions which obligatorily pied-pipe; (2) multi-word verbs and prepositions triggering obligatory stranding; and (3) antecedents in relative clauses causing obligatory pied-piping.

3.2.3.1 Obligatorily pied-piped prepositions

Scattered in several publications, one can come across the following set of prepositions, all of which are said to require obligatory pied-piping (Johansson and Geisler 1998: 75; Ross 1986: 134ff.; Ungerer *et al.* 1996: 206):

(3.106) temporal: *during*
location: *beyond*, *under*, *underneath*
means: *by means of*
contingency[17]: *because of*, *due to*, *owing to*, *in spite of*

Obviously, despite their different semantic function, all of the prepositions above are part of the 'group of less frequent, typically disyllabic and morphologically complex prepositions with a narrow range of more specific meanings and uses' (König and Kortmann 1991: 112). Now, frequent prepositions, like *on*, are often employed in phrasal verbs (e.g. *turn on*). Thus, since *on* surfaces regularly as part of complex lexical [V PRT] verbs, a reanalysis operation of [V] [P NP] to [V P] [NP] (see above) can reinterpret *on* fairly easily as part of a complex V. In contrast to this, there are (virtually) no complex lexical verbs including the infrequent prepositions in (3.106).[18] As such, the reanalysis operation lacks an overt model for creating a complex V like *kill by means of*.

Another factor leading to the obligatory pied-piping of the prepositions in (3.106) has to do with the prototypical function of the PPs which they project: *during*-PPs, for example, always function as temporal adjuncts (3.107), whereas *under*-PPs are often used as 'domain/respect' adjuncts, which 'identify a relevant point of reference in respect of which the clause concerned derives its truth value' (Quirk *et al.* 1985: 483; (3.108)):

(3.107) a. The lecture **during which** John finished his homework
b. *The lecture **which** John finished his homework **during**

(3.108) a. This page states the terms and conditions **under which** you may use the Website. (http://www.step.org.uk/terms.aspx, last accessed 27 June 2010)
b. * This page states the terms and conditions **which**you may use the Website **under**

Thus, the prepositions in (3.106) normally head sentence adjuncts, which, as pointed out earlier, do not add participants to an event, but modify it with respect to its temporal and physical location, its causes or truth value.[19] Since sentence adjuncts were identified as the most resistant with respect to stranding (section 3.2.2), the fact that some infrequent and complex Ps heading such adjuncts should obligatorily cause pied-piping should not be surprising.

17 Contingency adjuncts give information as to the cause, reason, purpose, result, etc. of an event (Quirk *et al.* 1985: 479).

18 As with any hard and fast rule, there are, of course, exceptions: take e.g. the idiomatic verbal complex *to bring under pressure*.

19 A potential exception being *by means of*, which specifies the means, i.e. 'abstract instruments', with which an action is accomplished: *The trick* ***by means of*** *which he obtained the money was a very old one* (from Ungerer *et al.* 1996: 206).

3.2.3.2 Obligatorily stranded prepositions

In contrast to the 'sentential adjunct' prepositions which obligatorily have to pied-pipe, there are also prepositions which have to strand obligatorily. These include prepositions which 'are part of long multi-word expressions such as transitive phrasal-prepositional verbs … e.g. *fob N off with*, *fix N up with*, *put N down to*, *let N in on*, *put N up to*' (Trotta 2000: 185):

(3.109) a. The old car **which** we **fobbed him off with**
b. *The old car **with which** we **fobbed him off** (from Trotta 2000: 185)

(3.110) a. **What** did you **put him up to**?
b. ***To what** did you **put him up**?

As Johansson and Geisler point out, the class of obligatorily stranding prepositions also includes 'highly idiomatic constructions such as the verb + particle + preposition in (*get*) *rid of*' (1998: 77):

(3.111) a. a … suite **which** we **got rid of** when we moved to this house (BNC: J8G.489; from Johansson and Geisler 1998: 77; emphasis added)
b. *a … suite **of which** we **got rid** when we moved to this house

Further complex idioms with obligatorily stranding prepositions include *make light of* (V-Adj-P), *let go of* (V-V-P), and *get wind of* (V-N-P) (cf. Ross 1986: 135).

Now, due to their non-compositional meaning, the above complex verbs can all be treated as 'multi-word verbs', i.e. as single lexicon entries. As such, the prepositions in these idioms will not have to be reanalysed since they are already part of a (lexical) complex verb. There are, however, also other instances where simple prepositions seem to induce obligatory preposition-stranding (see also Quirk *et al.* 1985: 664, 818, 1052):

(3.112) a. **What** does it taste **like**?
b. ***Like what** does it taste?

(3.113) a. **What** did you do that **for**?
b. ***For what** did you do that? (both examples from Trotta 2000: 63)

In a way the discontinuous *wh*- and P elements in the above examples appear to be lexically stored complex items: in (3.112) the *what … like* string has a meaning which can be paraphrased with *how* (cf. *How did it taste?*), while in (3.113) the unit *what … for* corresponds to a *why*-sentence (cf. *Why did you do that?*).

Finally, C.-J. Bailey points out another idiosyncratic effect concerning the copula *be*: he claims that stranding with *be* is obligatory in cases where pied-piping 'would result in a form of *be* standing at the end of the clause' (Bailey 1986: 165) such as *That's a topic *about which the book is* (example also from Bailey 1986: 165). The idiosyncrasy of this constraint is especially notable since *be* can occur at the end of clauses which do not involve preposition placement: cf. *That's where it is* (Bailey 1986: 165, fn. 6). Again this seems to require the postulation of a lexically stored constraint.

Thus, as the above discussion has illustrated, idiosyncratic lexical effects also have to be taken into account when discussing preposition placement in English.

3.2.3.3 Antecedents which induce obligatory pied-piping

Interestingly, in relative clauses, there are also a number of antecedent nouns which obligatorily demand preposition pied-piping: e.g. *way*, *manner*, *extent*, *degree*, *sense*, *point*, *time* and *moment* (see Johansson and Geisler 1998: 74; Ross 1986: 132):

(3.114) a. **the moment at which** the accident took place #from **the time at which** he looked over the shoulder (LLC: 12.4.<42–46, 284–5>)
b. there is **a point at which**, you know, he's always trying to evade sanctions, … (BNC: HM4.361)
c. The **degree to which** a school does or doesn't have immigrants is irrelevant (BIRM: GS0001)
(all examples from Johansson and Geisler 1998: 74; emphasis added)

(3.115) a. ***the moment which** the accident took place **at** #from ***the time which** he looked over the shoulder **at**
b. there is ***a point which**, you know, he's always trying to evade sanctions **at**, …
c. ***The degree which** a school does or doesn't have immigrants **to** is irrelevant

Example (3.115) shows that stranding a preposition in relative clauses modifying the above set of antecedents always produces a result of doubtful acceptability. Again, the obvious reason for this obligatory pied-piping tendency comes from the syntactic functions of the PPs: they are all sentential adjuncts either of manner (*way*, *manner*), degree (*extent*, *degree*), or respect/time (*point*, *time*, *moment*). At the same time, however, it will also have to be investigated whether a particular 'antecedent noun + P + *which*' sequence, such as, *way in which*, might also have been stored in the lexicon as a complex antecedent and relativizer structure. Such complex lexical items would then be on a par with free relativizers such as *what* (cf. 3.1.1) in that they encode head as well as relativizer properties.

As this chapter has tried to show, a simple dichotic complement–adjunct classification is insufficient for the syntactic description of preposition placement in English. For the empirical corpus study it was therefore decided to employ a more fine-graded classification based on the above findings, and incorporating Quirk *et al.*'s (1985: 479–86) classification of adjuncts (see Table 4.2 in section 4.1 for further information on the coding of the variable PP TYPES).

3.3 Level of formality

Besides the intralinguistic factors mentioned so far, an important extra-linguistic factor restricting preposition placement mentioned in virtually every publication on preposition-stranding is the level of formality: preposition-stranding is usually associated with speech and informal written contexts, whereas pied-piping is preferred in formal writing (see e.g. Biber *et al.* 1999: 107; Leech 1996: 375). This stylistic distribution of preposition-stranding has its foundations in the prescriptivist tradition of the eighteenth century:

The Preposition is often separated from the Relative which it governs, and joined to the verb at the end of the Sentence … as, 'Horace is an author, whom I am much delighted with' … This is an idiom which our language is strongly inclined to; it prevails in common conversation, and suits very well with the familiar style of 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. (Robert Lowth, *A Short Introduction to English Grammar*, 1762; cited in Aitchison 1991: 11)

The prescriptive grammarians' preference of preposition pied-piping over stranding resulted from their admiration for Latin, which 'was widely regarded as the most perfect of languages' (Aitchison 1991: 9) and therefore considered a model for the English language. Now, in Latin prepositions cannot strand but instead always have to precede their complements (cf. also the etymological source of *preposition*: Lat. *praeponere* 'put before'; Chalker and Weiner 1994: 310). Consequently, it is not surprising to find that Lowth calls pied-piping, which follows the Latin model, 'more graceful', while he would like to restrict the 'un-Latin' stranding of prepositions to 'common conversation' and 'the familiar style of writing'. Nevertheless, considering that 'Lowth's approach was essentially proscriptive' (Tieken-Boon van Ostade 2006: 544), the above statement is actually fairly lenient since it does not generally proscribe preposition-stranding. In this context, it is interesting to take a look at other prescriptive grammars from the eighteenth century.

As Alston (1965) points out, the most popular eighteenth-century grammars were Murray (1795), Lowth (1762), Ash (1760) and Fisher (1750) (cf. also Tieken-Boon van Ostade 2006: 542). Concerning preposition placement, Murray's view is almost a verbatim quote of Lowth's statement above:

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 their 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; bold emphasis added)

The first thing to note is that upon discussing relatives, Murray also includes the interrogative example *Whom wilt thou give it to?* vs *To whom wilt thou give it?*. As the quote shows, he furthermore agrees with Lowth as to pied-piping being more 'solemn and elevated' but stranding being common in conversation. The most notable difference is Murray's correction of Lowth's *This is an idiom* ***which*** *our language is strongly inclined* ***to*** to *This is an idiom* ***to which*** *our language is strongly inclined*. Concerning Lowth's stranded version Tieken-Boon van Ostade claimed that '[i]t is the kind of joke purists enjoy making, such as "a preposition is a word you cannot end a sentence with," or "this is something up with which I will not put."' (2006: 551). While that may have been the case, the joke seems to have been lost on Murray.

Moving on to the other two important grammarians of the time, it appears as if Ash (1760) did not comment on preposition placement, while Fisher (1750) explicitly argues that

> The *Preposition* is frequently transposed as, *who do you dine with*? For, with *whom do you dine*? [TH: italics sic!] *what Place do you come from*? For, *from what Place do you come*?
>
> Q. *May Words in Sentences be placed in what Order we please?*
>
> A. No; but we must in this, as well as in all other parts of *Grammar*, follow the Use of the best Speakers and Writers.
>
> ☞ The clearest and best Writers in *Prose* have the fewest *Transpositions* in their Discourses; and in *Poetry*, they are never used but when the nature and Harmony of the Verse require it (Fisher 1750: 123)

Thus, Fisher also acknowledges the fact that prepositions are often stranded (note how she considers this a case of 'transposition', i.e. for her pied-piping is the underlying normal structure). However, since the best speakers and writers predominantly favour pied-piping, people should try to follow their model (though Fisher also concedes that in poetry prepositions might need to be stranded because of metrical constraints). Even more striking is the fact that despite her preference for pied-piping she herself occasionally strands prepositions: '*What do you learn* Grammar *for*? … *What does* Grammar *treat of*?' (Fisher 1750: 1). Unlike Lowth's stranded example above, which might be interpreted as a linguistic joke, these two sentences appear on page 1 of Fisher's grammar, long before she covers preposition placement. Yet, both examples are interrogative clauses. Thus functional constraints (cf. 3.1.2) might have led to the choice of stranding, leaving the *wh*-word at the front of the clause to signal interrogative force. (In addition to this, the *what* … *for* sequence in the first example might already have acquired its status as a discontinuous lexical item meaning 'why'; cf. 3.2.3.2.)

Having surveyed the most important eighteenth-century grammars, it needs to be emphasized that these acknowledge stranding as a grammatical structure of spoken, informal English. Rather surprisingly, however, the prescriptive tradition emerging from these grammars then endorsed the slogan 'it is incorrect to end a sentence with a preposition' (Pullum and Huddleston 2002: 627). Pullum and Huddleston correctly point out that '[s]tranded prepositions often, but by no means always occur at the end of the sentence' (2002: 627). Nevertheless, lacking the metaphorical label 'stranded preposition', prescriptivists came up with the above hard-and-fast rule. The rule itself is attributed to Dryden, who criticized Ben Jonson's phrase *The bodies that those Souls were frighted from* (*Catiline*, 1611): 'The Preposition in the end of the sentence; a common fault with him and which I have but lately observ'd in my own writings' (John Dryden, *Defence of the Epilogue* 1672, ll. 55–76; taken from Görlach 1994: 202).

Usually it is implied that the simpler rule, i.e. the strict ban on stranded prepositions, has dominated the prescriptive tradition (cf. Sundby, Bjørge and Haugland 1991: 426–8; Pullum and Huddleston 2002: 627; Yáñez-Bouza 2006):

> Instead of being dismissed as unsupported foolishness, the unwarranted rule against stranding was repeated in prestigious grammars towards the end of the eighteenth century, and from the nineteenth century on it was widely taught in schools. The result is that older people with traditional educations and outlooks still tend to believe that stranding is some kind of mistake. (Pullum and Huddleston 2002: 627)

Yet, in addition to the fact that the most influential eighteenth-century grammarians were very well aware of the influence of the level of formality, Görlach also stresses the fact that there was a tendency in the nineteenth century to sever 'the ties that bound English to Latin' (1999: 76). The most prominent prescriptivist of the early twentieth century, Henry Fowler, even claimed that '[t]hose who lay down the universal principle that final prepositions are "inelegant" are unconsciously trying to deprive the English language of a valuable idiomatic resource, which has been used freely by all our greatest writers except those whose instinct for English idioms has been over powered by notions of correctness derived from Latin standards' (Fowler 1926: 458). Modern usage guides also reject the strict ban on stranded prepositions, calling the prescriptive rule 'one of the most persistent myths about prepositions in English' (Burchfield 1996: 617), which 'should be disregarded' (Weiner 1983: 166). Instead, as pointed out above, the choice between preposition-stranding and pied-piping with *wh*-pronouns is seen as a matter of style and obligatory stranding contexts such as passives are pointed out (e.g. Leech 1996: 375; Weiner 1983: 166).

Whenever there is a choice with respect to preposition placement, present-day usage guides thus emphasize that pied-piping is more or less restricted to formal registers. This has led some researchers to propose that in present-day English stranding is in fact the unmarked option, and that preposition pied-piping is only acquired, along with other formal writing skills, through formal education. Working on the acquisition of relative clauses, McDaniel, McKee and Bernstein (1998: 309), for example, claim that 'preposition pied-piping is not a natural option in English, but rather a prescriptive artifact probably picked up during schooling'.

As will be recalled, in section 3.1.2 it was argued that from a processing perspective relative clauses are probably the context which favours pied-piping most. Thus, for present-day English, McDaniel, McKee and Bernstein's study is of particular interest: could it really be that even in *wh*-relative clauses (for which Van den Eynden claims that 'stranding is not really an option' (1996: 444)), pied-piping is only acquired during formal education?

Investigating the relative clause production of 115 (American) children from the age of 3 to 11 with an elicitation test,[20] McDaniel, McKee and Bernstein found that even the youngest subjects already produced *that*+$P_{stranded}$ constructions like *that one that Baby Bob is jumping over*, but that none of their subjects employed the alternative P_{piped}+*wh*-pronoun option (indeed, their subjects only produced *that*-introduced relative clauses; see McDaniel, McKee and Bernstein 1998: 315). In a second set of experiments, the researchers then decided to test the subject's grammaticality judgements with regard to *wh*-relative clauses. For this, an experimenter would perform a short scene with toys, and then ask the subject whether a sentence related to the observed situation was grammatical; e.g. 'Is it right or wrong to say, "This is the robber who Dorothy talked to"? … "This is the robber to whom Dorothy talked"?' (McDaniel, McKee and Bernstein 1998: 318). As their results showed, virtually all subjects considered the stranded variant grammatical. In contrast to this, the pied-piped alternative enjoyed a relatively low grammatical status: only 3% of the youngest subjects (3–5 years old) and 6% of their 'Middle' group (6–8 years old) classified the pied-piped relative clauses as grammatical. Even within the oldest subject group (9–11 years) only 54% of the children considered the pied-piped alternative grammatical. On the other hand, 90% of an adult control group identified the pied-piped relative clauses as grammatical (cf. McDaniel, McKee and Bernstein 1998: 323).

All in all, McDaniel, McKee and Bernstein's study seems to imply that stranding, and not – as expected from cross-linguistic data (cf. section

[20] Basically, a first experimenter performed a story with a set of toys which the subject watched together with a second experimenter. Once the story was over, the second experimenter covered her eyes, and the first experimenter pointed to one of two identical toys, which could only be distinguished from each other by what had happened to them in the story. Then the child was asked to instruct the second experimenter to pick up the toy which the first experimenter had pointed to. Thus, children were forced to produce relative clauses like *the girl that the giraffe is sitting on* (McDaniel, McKee and Bernstein 1998: 313f.).

3.1.4) – pied-piping is the natural choice with *wh*-pronouns in present-day English. In the light of these results, it obviously appears as if pied-piping would not be part of the grammar originally acquired by children, but a prescriptive rule 'learnt' later via formal education.

However, taking a closer look at McDaniel, McKee and Bernstein's study reveals that it might not be possible to draw such far-reaching conclusions from their experiment. First of all, even though only a small number of their subjects considered the pied-piped variant grammatical, it follows that the grammar of this small group of children (e.g. 3% of the 3–5 year olds) already includes this structure.

Secondly, in general, the relatively low percentage of subjects accepting pied-piping might be due to the type of sentences used for the grammaticality judgements: the material only included complement PPs (*talk to*) but no sentential adjuncts. In fact, at no point in their article do McDaniel, McKee and Bernstein discuss the potentially different behaviour of different PP types. However, within their experiment it would have easily been possible to test sentences like *This is the house in which Dorothy met the robber* vs *This is the house which Dorothy met the robber in*, which might have produced a different result.

Finally, it might be that both pied-piping and the *wh*-relative pronouns are strongly associated with formal contexts (cf. section 3.1.2.2). Then it is to be expected that younger children simply have not encountered enough formal situations in which these phenomena occur frequently. While formal instruction in school is clearly such a situation, this need not be the only context. The increase in acceptability of pied-piping with increasing age might also have to do with the fact that older children have been to more church services or other formal occasions. It would definitely be surprising if explicit formal education on preposition placement should be the most important factor.

The above qualifications should be particularly true for Britain, where grammar teaching has undergone considerable changes over the past fifty years: up to the 1960s there was indeed a focus on explicit grammar teaching which was essentially prescriptive and assumed that 'there is a correct standard form of the language normally only found in writing' (Philp 2001: 724). From the early 1960s onwards, however, the focus has shifted towards more descriptive and usage-based grammatical descriptions (cf. Philp 2001: 727). The Cox Report (Department of Education and Science 1989), for example, emphasizes that the grammatical description of English taught in schools should be 'a form of grammar which can describe language in use [as well as be] able to describe the considerable differences between spoken and written English' (Department of Education and Science 1989: 4.28). Thus, generally speaking, British citizens who started their schooling in the 1960s or later (i.e. speakers of ages about 50 and under) should not have received the same kind of prescriptive, explicit grammar teaching as their predecessors.

Table 3.5 *PP types used for stranding and pied-piping with* wh-*relativizers in studies of written and spoken PdE (adapted from Bergh and Seppänen 2000: 307; supplemented by additional data from Johansson and Geisler 1998: 70)*

written corpora	stranding	%	pied-piping	%
Bengtsson (1996)	3	5	62	95
Van den Eynden (1996)	6	3	179[a]	97
Johansson and Geisler (1998) [LOB]	27	3	1053	97
Trotta (2000) [BROWN]	12	1	1054	99
total (written)	48	2	2348	98
spoken corpora	**stranding**	**%**	**pied-piping**	**%**
Quirk (1957)	18	17	86	83
Johansson and Geisler (1998) [LLC]	50	21	190	79
Johansson and Geisler (1998) [spoken BIRM]	129	14	799	86
Johansson and Geisler (1998) [spoken BNC]	65	31	147	69
total (spoken)	262	18	1222	82

[a] In Bergh and Seppänen's table, Van den Eynden is claimed to have found 166 piped tokens. The actual figure for piped WH-relative clauses, however, was found to be 179 (cf. Van Eynden's 1996: 442).

Consequently, for such speakers it seems unreasonable to assume that their us of pied-piping in *wh*-relative clauses can simply be accounted for by a taught prescriptive 'never end a sentence with a preposition' rule.

Nevertheless, if despite these objections, it actually turned out that children strongly favour stranding over pied-piping with *wh*-relative pronouns with all PP types, then it would obviously be interesting to see in what way formal education influences these preferences. Since modern usage guides tend to limit pied-piping to formal written registers, it should prove insightful to compare the results of studies on preposition stranding with *wh*-pronouns in written adult language, with studies examining spoken corpora. Bergh and Seppänen (2000: 306f.), for example, present an overview of five corpus studies[21] on preposition stranding in *wh*-relative clauses, presented in Table 3.5.

Upon comparing the two subtables, the most striking finding concerns the fact that preposition pied-piping is always preferred over stranding,

[21] The data analysed by these authors comprises the following corpora: Johansson and Geisler's study (1998) has already been mentioned in section 3.2.1. Bengtsson (1996) is an unpublished paper which investigates an 85,000-word corpus consisting of writings of G. B. Shaw; while Trotta examined relative clauses in a corpus of American written English (BROWN). Van Eynden's corpus consisted of texts from British newspapers (tabloids as well as quality newspapers), and Quirk analysed tape-recordings of English adults. All these studies – with the exception of Trotta, who looked at American Standard English – used corpora representative of Standard British English.

regardless of the particular mode: on average, 98 per cent of the prepositions are pied-piped in the written corpora and 82 per cent in the spoken corpora. If stranding were indeed the natural choice with *wh*-relative pronouns, then, in the light of Table 3.5, the influence of formal education would be immense: it would not only affect a speaker's formal written register, but would apparently install pied-piping as the default choice for both spoken and written language. Again, it seems doubtful that explicit prescriptive teaching should be responsible for such an across-the-board effect. Instead, it makes more sense to assume that pied-piping with *wh*-relative clauses is preferred due to processing constraints (cf. section 3.1.3). In addition to that, as pointed out in section 3.1.2.2, the *wh*-relative pronouns themselves are felt to be more formal than the *that*- and Ø-relativizers. Once speakers encounter more formal occasions, they will also come across more *wh*-pronouns, which favour pied-piping. This mutual co-occurrence of *wh*-relative pronouns and pied-piping thus explains the preposition placement preference in spoken and written registers. If pied-piping with *wh*-relative clauses really was nothing but a reflex of the prescriptive ban on stranding allegedly taught at schools, one would expect that in the written mode speakers have enough planning time for this rule to affect their language. Since the spoken channel leaves less time for conscious rules to interfere however, the strong preference for pied-piping in this mode (as indicated by the data in Table 3.5), does not lend itself to such an explanation. In the words of Halliday (1988: 38): 'Our ability to use [spontaneous spoken] language depends critically on our not being conscious of doing so – which is the truth that every language learner has to discover, and the contradiction from which every language learner has to escape.'

Besides, as Weiß notes, 'modern standard languages … are not only learned by instruction (e.g., in schools), but to a large extent acquired as L1 too' (2001: 94). Standard British English, for example, is not only acquired via formal education, it is also the sociolect associated with the middle and upper classes (cf. Leisi 1985: 191f.). As such, considering that adult speakers of Standard British English employ the '$\text{P}_{\text{pied-piped}}$+*wh*-pronoun' strategy quite frequently in spoken language (see Table 3.5), one might even expect social factors, such as class, to have a considerable effect on the acquisition of pied-piping.

Turning back to the data in Table 3.5, a closer look at the actual data of the individual studies reveals that the underlying distribution is again a lot more complex than it seems at first sight: as pointed out in section 3.2.1, all three of Johansson and Geisler's studies on stranding in spoken English confirmed the influence of the syntactic function of a PP. In their LLC data, for example, 61.8% (47/76) of prepositions in a complement PP were stranded. As can be seen in Table 3.5, this means that 94% (47/50) of all stranded prepositions in the LLC are the head of a prepositional complement, while only 6% are part of an adjunct PP (3/50).

Furthermore, comparing the figures in Table 3.5 more closely, it becomes apparent that there is a considerable difference between the results from written and spoken corpora: in the written data, the percentage of stranded prepositions is extremely low (ranging from 1% to 5%). In contrast to this, stranding appears to be a more viable option in spoken language (making up 14%–31%). Thus, the above data indicates a potential influence of stylistic factors on preposition-stranding. However, since not all written registers are more formal than spoken ones (take e.g. private correspondence, which cannot be said to be more formal than legal presentations; see Biber 1988: 47ff.; Nelson, Wallis and Aarts 2002: 6f.), a simple spoken/written dichotomy might not suffice to assess the precise extent of this influence. For the analysis of the ICE data it will therefore be necessary to classify data not only according to their mode (spoken/written), but also with regard to their level of formality.

The necessity of such a careful analysis of written as well as spoken data is probably best illustrated by the negative example set by Van den Eynden (1996): drawing only on written data, she ignores the six stranded prepositions in her corpus, and boldly claims:

> However, the same set of data shows that variability in preposition placement with Wh-relatives … is not just a matter of register. Diachronic, dialectal and standard English all show that stranding is not really an option with Wh-[relative pronouns]; not now and not in the past. (Van den Eynden 1996: 444)

Obviously, from a methodological point of view, Van den Eynden's interpretation of her data is utterly unacceptable. As Bergh and Seppänen point out, 'grammatical system[s] may contain both frequent and infrequent structures as fully normal parts' (2000: 313). Therefore, instead of ignoring infrequent counterexamples, one should rather attempt to incorporate them into one's theory. Apart from this, if Van den Eynden had also examined spoken data, say, the LLC, she might have had to acknowledge the existence of stranding with *wh*-relative pronouns, since it is much more difficult to attribute one-fifth of all tokens (cf. Table 3.5) in a corpus to something like 'performance errors'.

The above discussion has focused on relative clauses. Yet, pied-piping is generally considered the more formal option regardless of the particular clause type (for those clauses that allow for both preposition placement options; cf. Biber *et al.* 1999: 106; Pullum and Huddleston 2002: 627; Quirk *et al.* 1985: 817; Trotta 2000: 64). Trotta (2000: 64), for example, tried to find out whether the level of formality also had an effect on interrogative clauses. For this, he investigated more formal text types from the Brown corpus ('informative texts', i.e. the Brown corpus categories A to J) with the more informal ones ('imaginative texts', i.e. the Brown corpus categories K–R). His results can be seen in Table 3.6.

Table 3.6 *Preposition pied-piping and stranding in interrogative clauses from the BROWN corpus by text type (adapted from Trotta 2000: 64)*

P placement	formal	informal	Total
Preposition stranded	32	68	100
Preposition pied-piped	44	13	57
Total	76	81	157

Cells whose observed frequency is significantly lower than their expected frequency have been shaded light grey, while those with an observed frequency significantly higher than expected have been shaded dark grey.

As noted in section 3.1.3, Trotta does not subject his data to any statistical test. A simple chi-square test, however, indicates that the observed effect of the level of formality is significant ($\chi^2 = 27.909$, df = 1, $p << 0.001$ for a Pearson's chi-squared test with Yates's continuity correction for 2×2 tables). Interestingly, however, the Cramer's ϕ value of 0.435 for this chi-square test indicates that the association of the factors is not particularly strong. In addition to this, the effect of the level of formality seems to be more pronounced with the pied-piped tokens: the PIED-PIPED × INFORMATIVE TEXTS and the PIED-PIPED × IMAGINATIVE TEXTS cells both have a χ^2-value larger than 7, i.e. a probability of $p < 0.01$. In contrast to this, the STRANDED × INFORMATIVE TEXTS and the STRANDED × IMAGINATIVE TEXTS only have a probability of $p < 0.05$ (due to their χ^2-values of 9.76 and 9.15, respectively). In order to assess the precise nature of these effects, however, it will be necessary to carry out a multivariate analysis to preclude any confounding effects (such as the possibility that the formal tokens in Table 3.6 consisted mostly of more adjunct-like PPs, while the informal tokens included a larger number of more complement-like PPs).

In clause types which admit variable preposition placement it has been claimed that pied-piping is favoured in formal contexts, while informal contexts favour stranding. Considering the complex interaction of preposition placement, clause types and level of formality, the exact influence of these three individual factors requires a multivariate corpus study which recognizes various degrees of formality (ranging from more informal spoken texts to very formal written ones).

3.4 PPs embedded in NPs and AdjPs

A potential factor influencing the placement of the prepositions which many researchers tend to ignore is the type of phrase into which a PP is merged. In fact, whereas most corpus studies investigate the influence of the syntactic function of PPs which are merged to a verbal projection, i.e. complement and adjunct PPs in VP, only Trotta explicitly draws attention to the fact that the prepositions of PPs embedded in NPs and AdjPs might behave differently with regard to stranding and pied-piping (2000: 60ff., 184f.).

Before looking at Trotta's data more carefully, it must be pointed out that PPs in NPs and AdjPs have also been classified with respect to the complement–adjunct classification:

(3.116) a. a student [of physics] [with long hair]
b. *a student [with long hair] [of physics] (taken from Radford 1981: 98)
(3.117) a. fond [of Mary] [in some ways]
b. *fond [in some ways] [of Mary] (taken from Radford 1988: 244)

Due to the fact that they must normally precede other PPs (cf. (3.116), (3.117b)), *of physics* and *of Mary* are identified as complements of *student* and *fond*, respectively (see Radford 1981: 98; 1988: 244). For *of Mary* this analysis is furthermore supported by the fact that *fond* obligatorily subcategorizes for *of* (cf. **he is fond*), while *in some ways* is an optional constituent. In contrast to this, both PPs are optional in (3.116). Yet, in addition to the word-order effect in (3.116b), there seems to be a semantic difference between the two PPs: in the sentence *John is a student of physics* only one property is attributed to John, namely that he studies physics. In *John is a student with long hair*, however, he is attributed two properties: that he is a student and that he has long hair (cf. Radford 1981: 98). Thus, *with long hair* can be said to add extra, optional information in (3.116).

A first question which needs to be addressed now is whether preposition-stranding and pied-piping are both also possible with NP and AdjP-contained PPs. In his interrogative data from the Brown corpus, Trotta only found four AdjP-contained tokens (two of which were stranded and two of which were pied-piped; Trotta 2000: 61). In addition to this, his corpus only had one relevant NP-contained token:

(3.118) **Of what new course** could it mark the beginning
(BUC F14:16; cited in Trotta 2000: 61)

The low token size does not allow for any statistical tests, but it is nevertheless noteworthy that 96.8 per cent (152/157) of Trotta's interrogative tokens are VP-contained PPs. Recalling Hawkins's 'Valency Completeness' (Hawkins 1999: 278, 2004: 210–15) principle, this result is actually to be expected: the human processor prefers fillers which can be integrated upon processing the main subcategorizor (i.e. the main verb of a sentence, *mark* in (3.118)). Regardless of the placement of the preposition in (3.118), the processor will always have to look beyond the main subcategorizor to integrate the filler:

(3.119) a. **[Of what new course]**$_i$ could it mark **[the beginning$_i$]$_{NP}$**
b. **[What new course]**$_i$ could it mark [the beginning **[of$_i$]$_{PP}$**]$_{NP}$

In neither (3.119a) nor (3.119b) can the filler be integrated after processing the main verb *mark*. From a processing perspective both sentences are thus more complex than filler–gap structures involving VP-contained PPs. Furthermore, processing constraints predict that (3.119a) should be preferred over (3.119b) since in the latter case the integration of the filler is only possible once the stranded preposition is parsed, which itself is embedded in the NP *the beginning*.

The fact that NP-contained stranded prepositions are particularly sensitive to processing constraints has recently received experimental support. Cowart (1997: 15–18) showed that processing constraints affect the acceptability of NP-contained stranded prepositions: the greater the complexity of the NP, i.e. the number of semantic properties that have to be processed, the less acceptable a structure is judged. Thus stranded prepositions in indefinite NPs (cf. 3.120a) are judged significantly better than in definite NPs (3.120b), which in turn are judged better than NPs containing a possessive genitive NP (3.120c):

(3.120) a. Who did the Duchess sell **a** portrait of?
b. Who did the Duchess sell **the** portrait of?
c. Who did the Duchess sell **Max's** portrait of?
(taken from Cowart 1997: 15)

As Cowart proves experimentally (1997: 16f.), this effect only applies to NP-contained PPs. Thus if a PP is VP-contained in structures maximally similar to (3.120a) or (3.120c) (i.e. *Who did the Duchess* [*sell* [*a portrait*]$_{NP}$ [*to*]$_{PP}$]$_{VP}$*?* and *Who did the Duchess* [*sell* [*Max's portrait*]$_{NP}$ [*to*]$_{PP}$]$_{VP}$*?*) stranded prepositions do not lead to a decreased acceptability judgement. Only in NP-contained PPs, which are more difficult to process to start with, does an increased complexity lead to stranded prepositions being considered less acceptable.

This processing complexity perspective can also account for Trotta's claim that NP-contained PPs 'are less resistant to stranding when the noun phrase in which they are contained is in subject predicative function' (2000: 62):

(3.121) a. **Of which country** is he the ambassador?
b. **Which country** is he the ambassador **of**?

(3.122) a. ***Of which country** did you like the ambassador?
b. ??**Which country** did you like the ambassador **of**?
(all examples taken from Trotta 2000: 62)

Trotta (2000: 62) maintains that the sentences in (3.121) are clearly better than the ones in (3.122). Now, copula verbs such as *is* in (3.121) are 'semantically weak' (Trotta 2000: 62), and therefore easy to process. In such cases the subject predicative NP actually adds the main semantic information about the subject NP, and thus both pied-piping and stranding are acceptable. In (3.122), however, a full verb acts as the main subcategorizor and a greater amount of semantic information has to be processed upon encountering *like.* Consequently, if the gap for the filler is contained within the object NP (as in (3.122a)) or within a PP that is contained in the object NP (as in (3.122b)), the resulting structure becomes much more difficult to process and thus less acceptable.

Trotta's judgements of (3.122a) as completely ungrammatical '*' and (3.122b) as being of extremely low acceptability '??' deserve some attention: normally, pied-piping should be preferred in structures that are hard to process. Accordingly, (3.122a) should actually be more acceptable than (3.122b). A reason for Trotta's unexpected claim that the stranded version is in fact better than the pied-piped one might have to do with the fact that stranding is generally favoured in interrogative clauses (section 3.1.3). In contrast to this, in relative clauses, where pied-piping seems to be the preferred choice (section 3.1.3), the stranded version does not appear to be better than the pied-piped alternative:

(3.123) a. the country **of which** you like the ambassador
b. the country **which** you like the ambassador **of**

Moreover, the PP in question is complement-like (*the ambassador* [*of the Eastern European country*] [*with the red beard*] vs **the ambassador* [*with the red beard*] [*of the Eastern European country*]), which would also generally favour the stranded alternative in both (3.121) and (3.122).

The observation that semantically weak copulas license preposition-stranding more easily than full verbs also has repercussions for AdjP-contained PPs:

(3.124) a. **Of what** is he capable?
b. **What** is he capable **of**?

As illustrated by (3.124), adjectives which head PPs contained in AdjPs normally co-occur with copulas and provide the main semantic information. Consequently such PPs should appear in stranded as well as pied-piped constructions, with the latter being slightly preferred for processing reasons. Again, however, the sentences in (3.124) are interrogative clauses in which a complement PP is questioned, probably leading to a slight preference of (3.124b) over (3.124a). Since most AdjP-contained PPs seem to be more complement-like, this might explain Trotta's claim that adjectives are 'less restrictive as matrix predications in terms of stranding than their NP counterparts' (2000: 184).

The above discussion demonstrates that in order to factor out potential effects of NP- and AdjP-contained PPs it is also necessary to take a look at more than one clause type. In his analysis of the relative clause-tokens in the Brown corpus, Trotta, for example, found that prepositions were pied-piped in 99% of all cases (1,054 out of 1,066 tokens; see Table 3.5). Interestingly, all stranded prepositions in his corpus were part of prepositional verbs (12 out of 1,066 tokens; see Trotta 2000: 182). In contrast to this, all 53 AdjP-contained and all 56 NP-contained PPs in his relative clause data exhibit pied-piping (Trotta 2000: 184f.).

Now, taking a closer look at the PPs embedded in NPs and AdjPs, Trotta – relying on his native-speaker intuition – claims that for the majority of the examples in his corpus there would be 'nothing ungrammatical about a stranded preposition' (2000: 184). As such he concludes that the striking preference for pied-piping in his corpus might be due to 'the possible effect of prescriptive attitudes and/or editorial policies [since the Brown corpus only consists of written, published texts]' (2000: 184).

(3.125) These inwardly dramatic moments showed the kind of 'opera style' ***of which*** Beethoven was genuinely capable, but which did not take so kindly to the mechanics of staging
(BUC C12:71; taken from Trotta 2000: 184)

With respect to (3.125), Trotta argues that the alternative *the kind of 'opera style' which Beethoven was genuinely capable of* is perfectly grammatical. In fact, the obligatoriness test (cf. section 3.2.1) shows that *capable* is obligatorily subcategorized for the preposition *of*, cf. **he is capable*. As such, one would expect that *of* should easily be stranded in less formal registers.

Concerning NP-contained PPs, Trotta also claims that even though his data only contains pied-piped prepositions, an alternative with stranded preposition would be available for most of his examples. Thus, for the example in (3.126), he provides the grammatical alternative *to a school which he is denied entrance to.*

(3.126) Should Congress authorize the Attorney General to file suit to accomplish admission of a child to a school ***to which*** he is denied entrance
(BUC J48:76; taken from Trotta 2000: 185)

Since *entrance* with the meaning 'admission to a place' is (optionally) subcategorized for *to* (cf. *he was denied *entrance for a school/into a school*), pied-piping should in fact be theoretically possible in (3.126). Yet what all of the above findings clearly illustrate is the need for a multivariate statistical analysis of preposition placement that distinguishes significant factors from accidental ones: do AdjP-contained PPs generally strand prepositions more easily then NP-contained PPs, or are Trotta's results due to distributional dependence effects (i.e. do the factors CLAUSE TYPE and PP TYPE suffice to account for the data regardless of the XP in which the PP is contained)?

From a theoretical perspective it will also be necessary to pinpoint the exact effect of all the processing constraints discussed above: when do these yield clearly ungrammatical structures and when do they only result in slightly less acceptable strings? In his Performance-Grammar Correspondence Hypothesis (PGCH), Hawkins, for example, argues that '[g]rammars have conventionalized syntactic structures in proportion to the degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments' (2004: 3). Thus the empirical data will have to be carefully scrutinized for processing effects which can be considered grammaticalized.

One effect that might also be prone to processing factors is a special kind of preposition-stranding and pied-piping of NP-contained PPs, which separates these PPs from VP- or AdjP-embedded ones: if, for example, the NP *the reports* in the declarative sentence in (3.127) is relativized, then, as can be seen in (3.128), a speaker can actually choose from a wide range of possible alternatives for the resulting relative clause (adapted from Ross 1986: 121):

(3.127) The government prescribes the height of the lettering on the covers of the reports.

(3.128) a. Reports [[**which**]$_i$ the government prescribes the height of the lettering on the covers **of**$_i$] are invariably boring.
b. Reports [[**the covers of which**]$_i$ the government prescribes the height of the lettering **on**$_i$] are invariably boring.
c. Reports [[**the lettering on the covers of which**]$_i$ the government prescribes the height **of**$_i$] are invariably boring.
d. Reports [[**the height of the lettering on the covers of which**]$_i$ the government prescribes$_i$] are invariably boring.

As (3.128) shows, if a relativized PP is embedded in an NP which itself is part of a more complex NP, the *wh*-pronoun on its own can introduce the relative clause, leaving the preposition *of* stranded (see (3.128a)). Alternatively, the *wh*-pronoun can also pied-pipe one or more of the higher NPs along. Thus in (3.128b), *the covers of which* is placed at the front of the clause, while in (3.128d) the entire object NP is moved to the clause-initial position.

Now, in all the examples in (3.128), placing an NP in relative-clause initial position leaves a preposition stranded. In contrast to this, Ross claims that pied-piping the various prepositions produces only ungrammatical results:

(3.129) a. * Reports [[**of which**]$_i$ the government prescribes the height of the lettering on the covers$_i$] are invariably boring.
b. * Reports [[**on the covers of which**]$_i$ the government prescribes the height of the lettering$_i$] are invariably boring.
c. * Reports [[**of the lettering on the covers of which**]$_i$ the government prescribes the height$_i$] are invariably boring.

For Ross, the examples in (3.129) are ungrammatical due to a constraint which, in his dialect, prohibits the pied-piping of prepositions if they are embedded within an NP, and if fronting would cause them to end up adjacent to the antecedent NP (1986: 123). Put differently, he thinks that pied-piping in (3.129) is ungrammatical since instead of interpreting the PPs as introducing a relative clause, they will be analysed as postmodifiers of the antecedent.

However, this is not a global constraint, but rather seems to be a problem arising due to the complexity of the examples in (3.129). The LLC, for example, contains several examples in which a P-*wh*-pronoun complex has moved out of a dominating NP, ending up adjacent to the antecedent:

(3.130) a. as an investment [[**of which**] the share price is [an integral part]] (LLC 2 2:1<630f.>)
b. if he's going to play a proper role in the society [**of which**] he is [a part] (LLC 5 2: <938f.>)
c. one tower [**of which**] he gave me [the name] (LLC 4 2: <901f.>)

Now, considering that a *wh*-pronoun embedded within an NP can pied-pipe along not only a preposition (as in (3.130)), but also other syntactic objects in the NP which dominate it (cf.(3.128)), one might speculate whether this has consequences for the general pied-piping tendencies of NP-contained PPs: since most NPs will be far simpler than those in (3.128), the general tendency to induce upward percolation of the *wh*-features might lead to an increase of pied-piped prepositions with NP-contained PPs. However, it must be pointed out that the apparently uncontrolled upward percolation of the *wh*-features in (3.128) seems (partly) to be semantically motivated: moving an NP along to the front of the clause automatically has the effect of emphasizing/focusing it, since this position (the C-system in generative analyses) is also associated with the topicalization/focusing of elements (see Grewendorf 2002: 66ff.). As such, it is obviously *the covers* which is emphasized in (3.128b), whereas *the lettering* plays a more important part in (3.128c).

In addition to this, the more syntactic material is pied-piped to the front of the clause as filler the less complex the gap structure becomes: once the *wh*-item has been processed as introducing a relative clause, in (3.128a) the processor has to integrate *the government prescribes the height of the lettering on the covers of* (i.e. *wh*-item + 12 words = 13 words) until the appropriate gap site for this filler has been encountered. On the other hand, in (3.128d) all the processor has to parse after the *wh*-word is *the government prescribes* (i.e. *wh*-item + 3 words = 4 words). While the relative-clause-internal filler–gap identification process is obviously facilitated by pied-piping more and more material to the front of the clause, this increases the processing complexity of another important domain: the co-indexation domain of antecedent noun and *wh*-relativizer (see Hawkins 2004: 150). For the interpretation of relative clauses it is vital that the *wh*-relativizer follows the antecedent noun as closely as possible so that it is clear which semantic entity the relative clause modifies. If a simple preposition intervenes between antecedent and *wh*-relativizer, as in the normal cases of pied-piping, this does not seem to pose a great problem for the processor. However, in (3.128d) nine words are separating the antecedent noun from the *wh*-relativizer (*the height of the lettering on the covers of*; thus this domain comprises antecedent + 9 words + *wh*-relativizer = 11 words), while in (3.128a) the *wh*-relativizer is immediately adjacent to the antecedent (thus in this domain only the antecedent and the relativizer have to be processed = 2 words).

Prosodic cues will surely also play an important role during processing, but simply contrasting the number of intervening words per domain shows the structural nature of this trade-off: with respect to the filler–gap domain (3.128a) requires nine more words to be processed than (3.128d) (13 words (3.128a) – 4 words (3.128d) = 9). On the other hand, with respect to the antecedent–*wh*-relativizer co-indexation domain (3.128d) is also more complex than (3.128a) by nine words (11 words (3.128d) – 2 words (3.128a) = 9; for details on this calculation cf. Hawkins's 'Total Domain Differential' (2004: 120)). Trying to minimize both domains to the least number of words, the ideal structure would thus be (3.128c): there the filler–gap domain consists of seven words (*which the government prescribes the height of*) and the antecedent–*wh*-relativizer co-indexation domain eight words (*reports the lettering on the covers of which*). Due to the fact that even the optimal candidate has extremely complex filler–gap and antecedent–*wh*-relativizer co- indexation domains, structures such as (3.128) can easily be predicted to occur only very rarely in normal spoken discourse.

In addition to the above remarks, Davis and Dubinsky claim that the stranding of NP-embedded prepositions is actually restricted by a complex interaction of verb meaning, and the semantic N-P relationship (2003: 8ff.). Now, even though these authors are working on stranding (or, in their terminology, 'extraction') in interrogative clauses (3.131), their approach can also be illustrated in relative clauses (3.132):

(3.131) a. **Who**$_i$ did he read [a book **about**$_i$]$_{NP}$?
b. ***Who**$_i$ did he burn [a book **about**$_i$]$_{NP}$?

(3.132) a. the man **who**$_i$ he had read [a book **about**$_i$]$_{NP}$
b. *the man **who**$_i$ he had burnt [a book **about**$_i$]$_{NP}$

In all examples in (3.131/3.132) the *wh*-pronoun *who* is the filler for the gap headed by the stranded preposition *about*. Furthermore, in both sentences, the PP headed by *about* is embedded within the NP *a book* (thus the structure is similar to *I have read* [*a book about it*], which is not to be confused with *I have read about it*, in which the preposition *about* is directly licensed by the verb *read*). Yet, whereas extracting *who* out of the NP is unproblematic for (3.131a/3.132a), it leads to the ungrammaticality of (3.131b/3.132b). Yet, why should such an effect be observed in structurally similar sentences?

Looking at the pied-piped alternatives of (3.131/3.132) reveals that the ungrammaticality of (3.131b/3.132b) is not due to an unlicensed stranded preposition:

(3.133) a. [**About whom**]$_i$ did he read [a book$_i$]$_{NP}$?
b. *[**About whom**]$_i$ did he burn [a book$_i$]$_{NP}$?
(3.134) a. the man [**about whom**]$_i$ he had read [a book$_i$]$_{NP}$
b. *the man [**about whom**]$_i$ he had burnt [a book$_i$]$_{NP}$

As can be seen in (3.133b/3.134b), even pied-piping the preposition does not improve the grammaticality of these sentences. Therefore, it follows that the ungrammaticality of (3.131b/3.132b) has nothing to do with the stranded preposition, but arises from a more general constraint on relative and interrogative clauses: it is simply not possible to relativize or question the prepositional complement of a declarative sentence like *he burnt a book about Minimalism*.[22]

Furthermore, there is one environment in which extraction out of NPs seems to be banned in virtually all languages: if the NP is in subject position:

(3.135) a. The man **who**$_i$ John saw [a picture **of**$_i$]
b. The man [**of whom**]$_i$ John saw [a picture$_i$]
(3.136) a. *The man [**who**$_i$] [pictures **of**$_i$] are on the table (from Grewendorf 2002: 17)
b. The man [**of whom**]$_i$ [pictures$_i$] are on the table

As (3.135a) shows, extracting the *wh*-pronoun *who* out of [*a picture of*$_i$] is perfectly acceptable if the NP functions as an object. If the NP has the subject function, however, extracting *who* alone, and thus stranding the preposition in the NP, is ungrammatical (3.136a). Since cross-linguistic evidence shows that extraction out of subjects is generally an illegitimate operation, the subject position is also known as a 'strong island' from which syntactic objects cannot 'escape' (Grewendorf 2002: 17). Yet, in English, at least, it appears to be possible to extract PPs out of subjects, as (3.136b) shows, where *of whom* has pied-piped to the front of the clause.

Regardless of how generative analysis might try to save the claim that no syntactic material can be extracted out of the subject position,[23] for the

[22] Without going into details of how Davis and Dubinsky handle the different grammatical status of (3.131a/3.132a) and (3.131b/3.132b; 2003: 8ff.), it should suffice to point out that in (3.131b/3.132b) the event described actually concerns a concrete object, i.e. a book, which is being destroyed, whereas in (3.131a/3.132a) the reading event refers to the non-physical content of the book which is being read. As such, the content of the book is clearly a central element to a verb like *read*. On the other hand, the burning of a book does not affect its non-physical content. As such, it seems impossible to focus on the destruction of the physical object, while focusing its non-physical content by relativizing it.

[23] One possibility would be to assume that the entire subject *pictures of whom* has been moved from SpecT to SpecC with subsequent further raising of *of whom* to a higher SpecTopic or SpecFocus position.

current study it is only relevant that prepositions cannot be stranded in subject NPs. From a processing perspective this can again be explained by Hawkins's Valency Completeness principle (2004: 210): if the human processor prefers to integrate fillers upon encountering the main subcategorizor of a clause, then structures such as (3.136a) in which a filler must be identified with a gap even before the main subcategorizor has been parsed clearly lead to great processing difficulties and should thus be strongly dispreferred cross-linguistically.

Considering the above observations, it will be necessary to investigate whether the corpora reveal a different stranding/pied-piping tendency for VP-, AdjP- and NP-embedded PPs. In addition to this, any differences will have to be investigated as to the potential influence of processing factors.

3.5 Complexity

In the preceding sections, a great number of factors, such as the stranding preference of prepositional verbs or the pied-piping tendency of NP-embedded PPs, were attributed to processing constraints. In addition to this, Trotta puts forth the hypothesis that an increase in the complexity of a clause might favour pied-piping: 'the distance between landing and extraction site **may** exert a **small** influence on the fronting/stranding choice' (Trotta 2000: 187; emphasis added).

Trotta (2000: 188) gives the following examples, in which stranding the preposition *with* seems generally acceptable (3.137c), but becomes ungrammatical, or at least problematic, if the relative clause is complex (3.137b).

(3.137) a. But questions ***with which*** committee members taunted bankers appearing as witnesses left little doubt that they will recommend passage of it. (BUC A02:4)

b. * ?But questions ***which***committee members taunted bankers appearing as witnesses ***with*** left little doubt

c. But the questions ***which*** he taunted us ***with*** left little doubt (taken from Trotta 2000: 188)

As (3.137a) shows, complexity does not affect the grammaticality of a construction if the preposition is pied-piped. Trotta argues that especially for non-complement PPs, i.e. the class of sentential adjuncts, pied-piping becomes virtually obligatory if the distance between extraction site and *wh*-pronoun is too long. He attributes this to the fact that pied-piping ensures that the preposition and *wh*-pronoun are correctly analysed as a single constituent whose syntactic function can immediately be established. If the preposition is stranded, however, too much intervening syntactic material might prevent the discontinuous PP from being interpreted as a single constituent (Trotta 2000: 188). This view is shared by Johansson and Geisler (1998: 76), who argue that '[o]n the whole, stranding is disfavoured if the relative clause contains complex complementation in the VP'.

Table 3.7 *Preposition pied-piping and stranding in interrogative clauses in the BNC corpus by text type (adapted from Gries 2002: 232)*

P placement	written	spoken	Total
Preposition pied-piped	122	0	122
Preposition stranded	125	54	179
Total	247	54	301

In light of the above discussion, it does not come as a surprise that earlier studies (e.g. Deane 1992; Gries 2002) have sometimes claimed that preposition placement can be almost entirely accounted for by evoking processing-based explanations. Of particular interest in this context is Gries (2002), since it is the only other quantitative multivariate corpus study on preposition stranding that I am aware of. Next I will therefore take a closer look at the findings of this study, before outlining the factors encoding complexity used in the present corpus study.

3.5.1 Gries's multivariate corpus study (2002)

Using data from the British National Corpus (BNC), Gries (2002) investigated preposition-stranding and pied-piping in interrogative clauses and found the overall picture shown in Table 3.7. As this shows, Gries's data generally support the effects mentioned in sections 3.1 and 3.3: in interrogative clauses stranding is preferred over pied-piping (with 179/301 = 59% vs 122/391 = 41%) and written texts exhibit more pied-piped prepositions than spoken ones (cf. 122/247 = 49% pied-piped tokens in written texts vs no pied-piped tokens in spoken texts).

Yet, the above figures also indicate a severe problem with Gries's data: the BNC is a hundred times bigger than the ICE-GB corpus (one hundred million versus one million words). Nevertheless, the ICECUP search of ICE-GB yields more stranded interrogative tokens than Gries reports for the BNC (ICE-GB: 202 tokens versus BNC: 179 tokens). While Gries only provides vague information on how his data were obtained (merely stating that he 'used a concordance program to search the British National Corpus' (2002: 232)), it seems that he did not extract all relevant stranded tokens. Given that stranded prepositions are not tagged in the BNC, this probably would have been impossible anyway. Instead he had to resort to some kind of holistic search algorithm (perhaps looking for all prepositions followed by a full stop or comma). Such an approach, however, seriously affects the validity of his statistical analysis since his results are based on a highly skewed data set which in all likelihood ignored a great number of relevant stranded tokens. (The ratio of pied-piped ICE-GB:BNC tokens is more plausible at 23 to 122 tokens, but that still leaves the problem that there is no way of telling whether the subset of stranded tokens extracted by Gries adequately represents the entire population of stranded interrogative clauses in the BNC.)

Besides this, there is another problem with Gries's statistical analysis: the multivariate statistics he uses. He draws on linear discriminant analysis (LDA), despite the facts that his data do not meet a basic assumption of this test (namely normality, as he concedes in fn. 10, 2002: 240) and that at least half of his predictors are factors and not numerical variables (another important prerequisite of LDA; see Baayen 2008: 154).

Bearing in mind these shortcomings, it will nevertheless be interesting to compare the results of the present study to the ones Gries found for his BNC data.[24] Based on the LDA results, Gries argues that the two most important factors influencing preposition placement in his study are 'the processing effort associated with the two word orders [i.e. preposition-stranding and pied-piping] and the knowledge of prescriptive grammar rules' (2002: 239). With respect to the former, Gries's multivariate analysis actually identified two types of processing-related variables as significant: (1) a quantitative complexity metric (the more lexical material (measured in number of syllables) intervened between filler and gap, the more pied-piping was favoured) and (2) a qualitative, semantic one (Kluender's (1990) notion of semantic barrierhood, with increased barrierhood also favouring pied-piping).

As I will discuss in more detail below, for the present study it was also decided to employ a (slightly different) quantitative complexity metric. In contrast to this, the semantic barrierhood of the bridge structure (the material intervening between filler and gap) was investigated only indirectly. The main reason for this has partly to do with the complexity of this concept. In essence, Kluender (1990: 188) claims that 'open class, low frequency referentially specific constituents are … difficult to extract over [and] closed-class, high frequency, referentially non-specific constituents are relatively easy to extract over'. Now, since semantic barrierhood thus comprises three partially independent effects (lexical class, frequency and referentiality) I do not see how it could straightforwardly be operationalized as a single quantitative variable. Gries apparently employed such a single numerical variable for barrierhood, but provides no information on how this index was calculated (instead he simply states that '[b]arrierhood is an index for open/

[24] In particular since many of the variables of the present study correspond to ones investigated by Gries: thus, his variable spoken vs written MODALITY encodes the factor LEVEL OF FORMALITY and his factors VERB and PREPOSITIONAL SEMANTICS can be seen as correlates of the variable TYPE OF PP. There are, however, also a number of variables which Gries uses to measure COMPLEXITY (such as the syllabic LENGTH OF THE PREPOSITION and the FREQUENCY OF THE PREPOSITION) which were not tested in the present study. On top of the fact that these variables had emerged as statistically insignificant in Gries's multivariate study, this exclusion was also based on the fact that these factors strongly correlate and interact with the variable TYPE OF PP (cf. sections 3.2.3.1 and 3.2.3.2).

closed-class words and frequency' (2002: 239 fn. 3)). On top of that, the semantic barrierhood of the bridge structure is a general constraint on extraction, which – with the exception of the effect of preposition placement – should affect competing pied-piped and stranded structures alike: thus both *Who did she talk to?* and *To whom did she talk?* are identical with respect to the semantic barrierhood of the intervening material (*did she talk*). The only difference between the two concerns the fact that in the stranded alternative the filler is also extracted over the stranded preposition. Thus in the stranding/pied-piping alternation it is especially the barrierhood of the preposition that assumes a central role: as Kluender (1990: 189) points out, prepositions are more open-class-like if they are used referentially, i.e. if they head adjunct PPs. In contrast to this, prepositions that are subcategorized by a verb behave more like closed-class items. Thus, following Kluender's semantic barrierhood constraint, prepositions are easier to extract over (i.e. strand) if they are part of prepositional verbs than if they head an adjunct PP. As will have become apparent, for the present study semantic barrierhood can therefore be said to be encoded by the various functional properties of the affected PP and its relationship with the head of the phrase in which it is embedded (cf. the levels of the factor group PP TYPE in section 4.1).

3.5.2 Complexity-related and -specific factors

Complexity is obviously a multi-faceted concept, with the factor group PP TYPE only capturing one aspect. Another effect which can be attributed to processing complexity is the alleged pied-piping preference with NP-contained PPs: in cases where a preposition is stranded in an NP the processor has to look into a phrase which is embedded in another phrase, the VP, in order to relate the filler to the correct gap site. In such structures pied-piping should therefore be preferred. These kinds of effects were thus also inspected via the variable TYPE OF XP IN WHICH THE PP IS CONTAINED.

Besides this, the variable CLAUSE TYPE can also be argued to test processing complexity: in interrogative clauses the filler–gap identification process involves matching a quantifier which acts as filler and a gap that carries variable information (whose values are to be specified by the answer to the question; cf. Grewendorf 2002: 75; Huddleston, Pullum and Peterson 2002: 1071). In contrast to this, relative clauses require far more domains which have to be processed: in addition to the filler–gap identification domain, the processor needs to parse the co-indexation domain of antecedent noun and *wh*-relativizer (cf. section 3.4). Moreover, the entire relative clause as a domain must be processed as a sentential adjunct to the antecedent noun (see Hawkins 2004: 150). Finally, if the relative clause is restrictive it also constitutes the so-called 'Referential restriction' domain (Hawkins 2004: 150): since restrictive relative clauses are necessary for the identification of the reference of the antecedent NP this information must also be part of the parsing process.

All of these processing domains indicate that processing relative clauses is far more complex than parsing the other clause types which allow variable preposition placement. Consequently, if pied-piping is preferred in more complex structures (because of Hawkins's (2004) 'Avoid Competing Subcategorizors' and 'Valency Completeness' principles; see section 3.1.4), then relative clauses should favour this variant more than any other clause type. Note, however, that even within the group of relative clauses there are more and less complex structures: non-restrictive relative clauses are not needed for the identification of the reference of the antecedent NP. They are thus slightly less complicated than restrictive relative clauses (Hawkins 2004: 240–2), while still involving more processing cost than interrogative clauses. With respect to preposition placement, this hypothesis thus predicts that non-restrictive relative clauses should favour pied-piping more than interrogatives, but pied-piping should be even more preferred in restrictive relative clauses.

These findings obviously furthermore support the decision argued for in section 3.1.3 that the relative clause data should be subjected to an independent multivariate analysis (since the factor 'restrictiveness' is obviously meaningless for all other clause types). Since relative clauses are by definition the most complex clause type, it was furthermore decided to test these data for purely structural complexity effects. In order to systematically investigate the influence of structural complexity on preposition-stranding/pied-piping, the complexity of the ICE data was analysed by adapting Lu's parsing-orientated 'Mean Chunk Number' hypothesis (2002). Basically, Lu assumes that in order to reduce the number of units in the working memory, a parser will, whenever possible, combine smaller units into a single larger one, a so-called 'chunk'. Take the example in (3.138):

(3.138)	on	whom	I	think	I	had	some	designs	or	intentions
chunks	1	1	2	3	4	5	6	6	7	6

After encountering the preposition *on*, the parser has to store one chunk. Since the following *wh*-relativizer and the preposition can be grouped together into a complete PP, the parser still only has to retain a single chunk after processing *whom*. All the following elements up to *some* cannot be grouped together as chunks, so that after processing *some*, the parser already has six chunks in the working memory. In the following string of words, only the co-ordinating *or* requires an extra chunk to be stored, while the others, i.e. *design* and *intentions*, can always be grouped into a noun phrase with the preceding chunk (i.e. *some design* and *some design or intention*). Adding up the number of chunks which the parser has to store at different times during the processing gives the 'Instant Chunk Number (or ICN)': 1+1+2+3+4+5+6+6+7+6 = 41 in (3.138). In a next step, the ICN is then divided by the number of words that had to be integrated. This formula gives a sentence's 'Mean Chunk Number' (or MCN): [ICN]/[Σ_{words}] = 41/10 = 4.1 for (3.138).

As can be seen in (3.139), the MCN of stranded prepositions is higher and thus assumed to be more complex than pied-piping:

(3.139)	whom	I	think	I	had	some	designs	or	intentions	on
chunks	1	2	3	4	5	6	6	7	6	7

Since the stranded preposition in (3.138) is identified as an extra chunk, the MCN of this sentence would be 4.7 [47/10].

In contrast to other quantitative measures of complexity, such as for example Gries's (2002) syllabic length of the bridge structure or Hawkins's (1994) IC-to-Non-IC value,[25] the MCN calculation thus implicitly encodes the hypothesis that pied-piping is generally less complex than stranding. While this was the reason why this measurement of complexity was chosen for the present study, it also required the following adaptation: simply calculating MCNs, all clauses with pied-piping would as a result have been categorized as less complex, and the effect of the underlying complexity would be lost. In order to overcome this problem, it was decided to reconstruct the base position of pied-piped prepositions, and to always take the MCN of the stranded alternative as a measure for the complexity of a construction. Consequently, the MCN for the pied-piped alternative in (3.138) was assumed to be 4.7, and not 4.1.

The present study acknowledges that processing complexity is a concept that has many different facets. In the empirical part complexity will therefore not only be measured by the variable COMPLEXITY but also by several of the other independent factors (PP TYPES, TYPE OF XP IN WHICH THE PP IS CONTAINED and CLAUSE TYPE).

3.6 Second-language effects: Kenyan English

The final question of the present study is whether L2 Kenyan English displays any effects of preposition placement different to L1 British English. In order to adequately address this issue it obviously first becomes important to get an overview of L2-specific properties of language learning before taking a closer look at English in Kenya.

25 Hawkins basically claims that a good indicator of structural complexity is the ratio of immediate constituents that have to be processed divided by the number of words encountered during processing. Take, for example, the two competing structures *the man* [*on* [*whom*]$_{NP}$]$_{PP}$ [*I*]$_{NP}$ [*relied*]$_{VP}$ and [*whom*]$_{NP}$ [*I*]$_{NP}$ [*relied*]$_{VP}$ [*on*]$_{PP}$. In both cases four ICs (two NPs, a PP and the VP) are realized by four words yielding an IC-to-Non-IC ratio of 4/4 = 100%. In contrast to this the MCNs for these structures would be [1+1+2+3]/4= 1.75 and [1+2+3+4]/4 = 2.5.

3.6.1 L2 language learning

Second-language acquisition/learning is 'the way in which people learn a language other than their mother tongue, inside or outside of a classroom' (R. Ellis 1998: 3). One of the characteristics of this acquisition process is the fact that the linguistic output of second-language learners displays far greater heterogeneity than that of first-language learners (both on an inter-speaker and an intraspeaker level). This variable output produced by second-language learners depends partly on situational factors (e.g. whether the language was acquired in natural settings or in a classroom): while learning languages in a natural setting seems to especially improve oral fluency, the advantage of classroom learning appears to lie in a greater increase in lexico-grammatical knowledge (Mackey 2006: 449).

In addition to this, individual differences stemming from the various L1 languages of the learners, their age and even factors such as motivation affect the learner's output. The effect of the learner's first language is also known as transfer. In cases where L1 influence leads to errors in a learner's L2 output, one also speaks of 'negative transfer' or 'interference' (see Ellis 1998: 51–4; Mackey 2006: 446). While research has shown that L1 interference cannot account for all L2 errors, it is also well known that partly conscious, positive transfer of L1 features does occur. Besides L1 transfer, age is usually considered of crucial importance: L1 acquisition is a fairly uniform and fast process, while L2 acquisition, on the other hand, is crucially affected by age: younger language learners are far more successful at learning second languages than adults (see e.g. Mackey 2006: 446). A point to note in this context is the distinction 'between the *rate of attainment* and the *ultimate level of success* [... since] even though adults and adolescents may initially be faster at learning an L2, children generally out-perform them in the long run' (Mackey 2006: 447). In addition to this, adults and older children, due to their cognitive capacities, will benefit more from classroom situations than younger children, who do better in natural settings (see Steinberg and Sciarini 2006: 123–35). Nevertheless, even though a sensitive period up to about puberty therefore appears to exist before which language acquisition in general is easier, occasionally even adult L2 learners achieve native-like proficiency (cf. Mackey 2006: 447). Finally, inter-individual differences can also be attributed to differences with respect to the motivation of language learners (for further potential factors, see Mackey 2006: 446–9). Motivation obviously includes integrative aspects (whether the learner perceives the target language group as positive and wants to associate himself with it) as well as instrumental aspects (whether learning an L2 is potentially associated with social or economic rewards; cf. Mackey 2006: 448).

Another integral component of second language acquisition is, of course, the input which a learner encounters. What is particular noteworthy is that input does not only consist of native-speaker data (with its underlying target-language properties), but also data produced by other non-native speakers, such as teachers or other learners. Thus learners do not only come across native-speaker models. In fact, especially learners not living in a country where the target language is spoken as the mother tongue of a number of speakers will probably have more contact with L2 variants of the target language (see R. Ellis 1995: 276).

Concerning the complex interaction of all the above components it must be pointed out that both situational factors and learner differences can be said to influence learner processes such as different learning, production and communication strategies (see R. Ellis 1998: 76–8): learners employ different communication strategies in natural settings and classroom situations, and learners of different ages, for example, have different cognitive strategies available to them (see R. Ellis 1995: 275f.). Moreover, learner differences and situational factors also affect the input available to the learner: motivation and personality clearly 'determine the quantity and quality of the input' (R. Ellis 1995: 275f.), and natural input is obviously different to data which has been specifically collected and prepared for classroom settings (see R. Ellis 1995: 275f.). Lastly, '[i]nput constitutes the data upon which the learner strategies work', but the input that a learner receives also depends on his or her communication strategies (R. Ellis 1995: 276).

Despite the fact that L2 output is highly variable, however, it has also been shown that all learners seem to acquire a second language in similar ways: 'For example, regardless of their L1 and the type of input they receive, learners of English pass through similar sequences of developmental stages when learning negation and question formation' (Mackey 2006: 451). One explanation of this phenomenon has been the concept of markedness, which '… refers to the general idea that some structures are more "natural" or "basic" than other structures' (R. Ellis 1998: 70). The precise definition of markedness differs from framework to framework: Chomskyan generative linguists claim that unmarked structures are those that arise from an innate grammar device, i.e. Universal Grammar, which can guide the acquisition of first and second languages (see R. Hawkins 2001; White 2003). In contrast to this, typological linguists only consider the frequency of a structure in the world's languages as a measure of markedness. While typological statements per se do not seem to be explanatory in nature, John A. Hawkins's approach to processing constraints (1999; 2004) actually seeks to explain the distribution of structures across languages: those structures which are easier to parse will also become grammaticalized in a greater number of languages, which in turn means that they are more unmarked (cf. also Eckman 2004; R. Ellis 1998: 70). Now, in section 3.1.4 it was pointed out that preposition-stranding is very rare cross-linguistically. As such, stranding would have to be considered as the marked, and pied-piping as the unmarked structure. It next becomes necessary to see whether this factor or any of the others discussed in this chapter influence the acquisition of preposition placement by L2 learners of English.

3.6.2 General properties of L2 preposition placement

The earliest study that touched upon the question of markedness and the acquisition of preposition placement by L2 learners of English was Mazurkewich's (1985) investigation of the acquisition of dative *wh*-questions (which, among other things, tested the structures such as *Who did you give the book to?* and *To whom did you give the book?*). Since pied-piping is taken to be the unmarked structure, Mazurkewich predicted that regardless of their first language, learners should always acquire the umarked alternative (i.e. pied-piping) before the marked one (i.e. stranding). Yet, testing two groups of learners, native speakers of Quebec French and native speakers of Inuktitut (an Inuit language), only the French learners showed the expected effects: they acquired preposition pied-piping before stranding, and even when they had acquired stranding they always produced more instances of pied-piping. In contrast to this, even during the early stages the Inuktitut speakers produced more stranded prepositions than pied-piped ones. As Bardovi-Harlig (1987: 388f.) correctly points out, this seems to be a case which can be partly accounted for by positive L1 transfer: Quebec French has obligatory pied-piping (but cf. Bardovi-Harlig 1987: 389) and thus it is not surprising to see that the French learners produce far more instances of pied-piping.

Inuktitut, on the other hand, has no prepositions, which precludes the effect of any L1 transfer. If unmarked structures should be acquired earlier by learners then one would expect the Inuktitut learners to learn pied-piping first and to exhibit marked preposition-stranding only later. Yet, Bardovi-Harlig (1987: 388–90) claims that the Inuktitut data suggest that markedness is not the only factor that influences the acquisition of grammatical features. Another factor which might assume importance is the frequency of a phenomenon in the target language, also known as the salience of a feature (Bardovi-Harlig 1987: 401). As mentioned in section 3.1.3, pied-piping in interrogative clauses is considered to be strongly dispreferred. Consequently, stranding in this context is far more frequent, and therefore more salient. The Inuktitut data thus imply that high salience in the target language can also lead to a marked structure being acquired earlier than its unmarked counterpart.

Bardovi-Harlig (1987) tested this claim with a group of ninety-five second-language learners with various L1s in an elicitation task. In the test, subjects were given sentences like (3.140) and (3.141):

(3.140) Peter threw a football *to Philip*

(3.141) The person ______________________ was Louise.
Allen lent \$100 to the person
(examples from Bardovi-Harlig 1987: 392)

Subjects were then asked to convert sentences like (3.140) into *wh*-interrogative clauses by questioning the italicized element (resulting in either *Who did Peter throw a football to?* or *To whom did Peter throw a football?*). Similarly, subjects had to combine the two sentences in (3.141) into a single sentence by turning the sentence below the line into a *wh*-relative clause (yielding either *The person to whom Allen lent $100 was Louise* or *The person who Allen lent $100 to was Louise*; cf. Bardovi-Harlig 1987: 391–2).

Bardovi-Harlig thus tested the acquisition of preposition placement in *wh*-interrogative and *wh*-relative clauses. Interestingly, for both these clause types learners first exhibited a third alternative: using no preposition at all (Bardovi-Harlig 1987: 393):

(3.142) Who did Susan create a costume?
(3.143) The policeman Bill reported the accident arrested him.
(examples from Bardovi-Harlig 1987: 393)

Establishing a dependency between a filler and a gap which is embedded in a PP involves a lot of processing cost (see sections 3.1.4 and 3.4). It is thus not surprising that learners should simplify such structures by simply omitting prepositions which introduce thematic participants. By omitting *for* in (3.142), the filler *who* becomes integrated by the learner's processor upon encountering the main subcategorizor of the clause *create* (*to create a costume for someone* thus becomes *to create someone a costume*). Similarly in (3.143) *to report the accident to someone* is simplified to *to report someone the accident* (note that, despite the fact that Bardovi-Harlig (1987: 393) allegedly only asked for *wh*-relative clauses, (3.143) contains a Ø-relativizer, which would have induced obligatory stranding). As several studies have shown (cf. e.g. Dekydtspotter, Sprouse and Anderson 1998; Kao 2001; Klein 2001), the omission of prepositions (also known as 'null prepositions') is a recurrent feature of L2 varieties.

In Bardovi-Harlig's data, sentences with no preposition continuously decrease with increasing learner proficiency. In addition to that, all learners first acquire preposition-stranding before pied-piping in both interrogative and relative clauses (see Bardovi-Harlig 1987: 394f.). Again this indicates that the salience of preposition-stranding in the target language seems to be more important than the fact that it is a marked structure. Interestingly, a small number of subjects sometimes 'employed a transitional strategy when moving from preposition stranding to pied-piping' (Bardovi-Harlig 1987: 399), in which preposition-doubling occurs (the preposition appears pied-piped as well as stranded):

(3.144) To whom did Allen lend a dollar to? (from Bardovi-Harlig 1987: 399)

The later learners in Bardovi-Harlig's study furthermore displayed a remarkable interaction effect: while pied-piping became the preferred choice for questions, stranding remained the favourite variant in relative clauses. This result is somewhat unexpected since relative clauses are considered more complex and accordingly more likely to favour pied-piping (see sections 3.1.3 and 3.5). Bardovi-Harlig explains these findings by pointing out that due to their complexity subjects employed the simplified 'no preposition' alternative much more often and much longer in relative clauses than in interrogative clauses (1987: 389). Moreover, preposition-stranding in relative clauses might also be more salient than previously thought, if learners take the many instances of stranded prepositions in *that*-/ Ø-relative clauses as a model for their *wh*-relative clauses (Bardovi-Harlig 1987: 402).

Kao (2001) tested Bardovi-Harlig's claims in a grammaticality judgement experiment involving ninety-nine Japanese university learners of English. In the experiment subjects had to judge whether they considered a sentence grammatical. If they found an item ungrammatical, they were asked to correct the sentence. Example (3.145) gives an overview of the factors investigated:

(3.145) a. *John *lived* that house two years ago.
b. *Which house did John *live* two years ago?
c. *This is the house which John *lived* two years ago.
(examples taken from Kao 2001: 200)

As (3.145) shows, Kao tested whether the Japanese learners detected the missing preposition *in* in these sentences, and if they did, whether they would strand or pied-pipe it in interrogative (3.145b) and relative (3.145c) clauses. In addition, it was also attempted to exclude a confounding effect that might have distorted Bardovi-Harlig's results: by collecting the subjects' responses to declarative sentences such as (3.145a), Kao was able to include only the judgements of participants who had already learnt that *live* and *in* form a collocational unit in English.

Kao's results corroborate Bardovi-Harlig's claim that the salience of preposition-stranding in English accelerates its acquisition by second-language learners: learners of all proficiency levels in the study favoured preposition stranding over pied-piping. In contrast to Bardovi-Harlig's study, however, subjects favoured stranding in both clause types with a frequency of 70% or over. The only effect observable of clause type was the expected one of pied-piping being slightly more frequent in relative clauses (with ratios of 16.5% null prepositions, 71.9% stranding, 11.7% pied-piping for learners with the lowest level of proficiency and 8.3% null prepositions, 77.2% stranding, 14.5% pied-piping for the highest level of proficiency) than in interrogative clauses (with ratios of 15.6% null prepositions, 80.5% stranding, 3.9% pied-piping for learners with the lowest level of proficiency and 8.3% null prepositions, 90.4% stranding, 1.3% pied-piping for the highest level of proficiency; Kao 2001: 201f.).

Other studies on the acquisition of preposition placement also supported the salience hypothesis: Sadighi, Parhizgar and Saadat (2004) working on the language of Iranian learners as well as Rezai's (2006) study of preposition placement by Persian-speakers showed that learners acquire preposition-stranding before pied-piping. Again, pied-piping is obligatory in both Iranian and Persian, but due to the salience of preposition-stranding in the L2 input, learners first seem to favour the more marked alternative. Only when they can be sure that pied-piping is also a grammatical option do they seem to use it more frequently (then potentially also due to positive L1 transfer).

All of the studies on the acquisition of preposition placement thus suggest that, as a result of its salience, preposition-stranding is first acquired and preferred by second-language learners of English. It needs to be pointed out that one reason for this result is the fact that all the studies above focused on the placement of prepositions which are lexically stored, i.e. prepositional verbs. Whether the same effects should be expected for more adjunct-like PPs (cf. section 3.2) remains doubtful, however, and requires further testing.

3.6.3 English in Kenya: Focus on preposition placement

So far, no empirical study on preposition placement in L2 Kenyan English has been carried out. In the light of the findings discussed in sections 3.6.1 and 3.6.2 it will therefore become necessary to take a closer look at the linguistic situation in Kenya and at the various factors affecting the acquisition of English.

Schmied (1991b: 52–7, 2004a: 924–5) argues that there are basically four factors that can be said to affect L2 varieties of English:

(a) general language-learning strategies
(b) L1 transfer
(c) exposure to written language
(d) influence of native-speaker models

By general language-learning strategies Schmied means universal psycholinguistic processes that affect second-language learning (cf. Skandera 2003: 25). These include simplification and overgeneralization strategies. One well-known feature of Kenyan English which displays signs of simplification is the preposition system. Prepositions are very often omitted in Kenyan English 'when they are "obvious" anyway (e.g. *put* [*in*], *protest* [*against*])' (Schmied 1991b: 68, 2004b: 931; sometimes this process also affects the particles of phrasal verbs, cf. Hancock and Angogo 1982: 316). As the previous section indicated, this phenomenon of null prepositions is fairly frequent in the L2 Englishes of speakers with various L1s. In addition to this, Kenyan English also shows 'a tendency to substitute *in* for *on*, *at*, or *to*' (Schneider 2007: 196; cf. Mwangi 2004). This process of extending the meaning of a single preposition is of course a general learning strategy.

Besides, another reason for this simplification might be the system of the local languages: Kiswahili, for example, has a very simple preposition system in which e.g. *mwituni* can mean 'at, to, in/inside, by/near/next to and from the forest' (Schmied 1991b: 68). Moreover, Kenyan English exhibits verb–preposition collocations not found in L1 varieties, e.g. *attach with* instead of *attach to*, or *concentrate with* instead of *concentrate on* (cf. Mwangi 2004; Schneider 2007: 196).

Other concepts subsumed under general language-learning strategies include markedness (Schmied 1991b: 52) and salience effects. In the light of the second-language studies on preposition placement mentioned in section 3.6.2, it can be expected that the salience of preposition stranding should also lead to its frequent use in Kenyan English, despite its typologically marked status. Furthermore, it is possible that null prepositions will occasionally surface or that prepositions other than expected from L1 varieties will appear from time to time.

With respect to L1 transfer, Schmied argues that the influence of negative transfer has been somewhat overestimated (1991b: 53). However, as discussed in section 3.6.1, several studies have indicated that it is especially positive, partly conscious, transfer that can affect second languages. It therefore becomes necessary to take a closer look at preposition placement in the local Kenyan languages.

Kenya is home to more than seventy tribal groups (Parkinson, Philips and Gourlay 2006: 43). While this figure is somewhat vague, a definite number cannot be given since

> distinctions between many groups are becoming increasingly blurred, largely as a result of migration to the cities and encroaching Western cultural values. Many smaller tribes have also come in under the umbrella of larger tribal groups to gain protection in intertribal disputes. (Parkinson, Philips and Gourlay 2006: 43)

All of these tribal groups speak some kind of vernacular, but it is not always easy to say whether the varieties of two tribes should be classified as two dialects of one language or two separate languages (Heine and Möhlig 1980: 9). Thus Mbaabu's figure of over forty indigenous languages (1996: 147) should only be considered a rough estimate.

Apart from a small percentage of Indian languages (mainly Hindi and Urdu) spoken natively by the Indian community, all of the vernacular Kenyan languages belong to either the Bantu, the Nilotic or the Cushitic language families: most Kenyans (about 65 per cent) speak a Bantu language such as Kikuyu (also known as Gikuyu), Kamba or Luyia. The second largest group (about 30 per cent) speak a Nilotic language such as (Dho)Luo or Kalenjin, while only 3 per cent of the population speak a Cushitic language like Boni (cf. Heine and Möhlig 1980: 10–55; Musau 2004: 60).

Interestingly, Swahili (also referred to as Kiswahili), which is the second official language of Kenya – and the second language of more than 50 per cent of Kenyans (cf. Heine and Möhlig 1980: 61) – is also a Bantu language. It is not a 'pure' Bantu language, however, since it 'has been in intensive contact with non-Bantu languages, most notably Arabic and in more recent times English' (Krifka 1995: 1397). Nevertheless, the 'influx of these languages … is largely confined to the lexicon; as far as syntax and morphology is concerned, Swahili can be considered as a fairly typical Bantu language' (Krifka 1995: 1397). Next I will therefore turn to preposition placement in Swahili.

As the linguistic literature on the topic reveals (see Barrett-Keach 1985; Brauner and Bantu 1967; Vitale 1981), Swahili does not license stranded prepositions: *wh*-question words remain *in situ* and generally do not appear clause-initially (see Krifka 1995: 1415 for details and possible innovations). In relative clauses, if the complement of a preposition is relativized, the preposition can not be stranded (3.146a) but must be followed by a pronominal clitic (3.146b):

(3.146)	a.	*mtu	amba-**ye**	watoto	wa-li-pony-w-a	na	ni	mganga
		man	PRO-REL	children	they-PST-cure-PASS	by	is	medicine man
		'the man who the children were cured by is the medicine man'						
	b.	mtu	amba-**ye**	watoto	wa-li-pony-w-a	na-**ye**	ni	mganga
		man	PRO-REL	children	they-PST-cure-PASS	by-**REL**	is	medicine man
		'the man who the children were cured by-him is the medicine man'						
		(examples from Vitale 1981: 96)						

Data such as (3.146) might lead one to predict that preposition-stranding in Kenyan English might sometimes be avoided by the insertion of a resumptive pronoun (e.g. *the man **who** I talked to **him***). The occurrence of such structures in Kenyan English could then be attributed to L1 transfer. Note, however, that resumptive pronouns facilitate the processing of relative clauses since they overtly indicate the gap site that a filler must be associated with. It is therefore not surprising that a resumptive (also known as 'shadow') pronoun strategy surfaces in a large number of varieties of English all over the world: it has also been reported for such distinct varieties as Scottish and Irish English (Miller 2004: 62 and Filppula 2004: 85, respectively), Gullah (Mufwene 2004: 364), Jamaican Creole (Patrick 2004: 427) and Black South African English (Mesthrie 2004: 967).

Turning to vernacular Bantu mother tongues, these – like Swahili – also do not allow preposition-stranding: neither Kikuyu (Leaky 1959) nor Kamba (Whitely and Muli 1962) nor Luyia (Appleby 1961; Donohew 1962) license stranded prepositions. This is possibly due to the fact that prepositional meanings are usually expressed by clitics. In Kamba, for example, the prefix *na-* means 'with' and combines with the interrogative word *ũũ* 'who' to give a complex question word *naũ* / *naũũ* 'with whom', which usually appears at the end of a sentence (Whitely and Muli 1962: 98). Such examples might potentially be seen as cases of pied-piping, which would mean that L1 transfer should favour this preposition placement variant. In relative clauses, however, pied-piping does not seem possible. Instead, in the Luyia sentence in (3.147), just like in the Swahili example above, a pronominal clitic appears behind a preposition which would otherwise be stranded:

(3.147)	omwana	owa	endi	**nina-ye**	alwala
	the child	**whom**	I-am	**with-he/she**	is ill

‘The child who I am with-it is ill’
(adapted from Donohew 1962: 18)

In contrast to the Bantu languages, at least one of the Nilotic languages allows preposition stranding: as Omondi (1982) showed, Dholuo has *in situ* P+*wh*-word structures in questions (*ni´ ang'ó* in (3.148)), but in relative clauses prepositions can be stranded (*.e* in (3.149)):

(3.148)	Ibîro	**ni´**	**ang'ó?**
	you-have-come	**for**	**what**

‘What did you come for?’
(adapted from Omondi 1982: 142)

(3.149)	puodhó	**má**	néně	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)

Thus there are at least some Kenyans for whom the salience of the input as well as positive transfer from their mother tongue should coincide in favouring preposition-stranding.

Finally, there is no information on preposition placement in Cushitic languages. However, since Heine points out that prepositional meanings in Boni are also expressed by clitics (1982: 54, 67), it is presumed that these languages behave similarly to the Bantu languages with respect to preposition placement (i.e. avoiding preposition-stranding by resumptive clitics).

While L1 transfer thus might have different effects depending on the mother tongue of the speaker, the exposure to written language should be a factor that should affect all Kenyans alike. Schmied claims that in African societies the written word carries more weight than the spoken form. As a result, ‘African speakers of English tend to reproduce characteristics of written English even in the spoken form. Grammatical constructions and lexical items from relatively formal registers or spelling pronunciations will often be used’ (1991b: 53; cf. also Abdulaziz 1991). This hypothesis is not uncontroversial: Hancock and Angogo, for example, maintain that ‘East African English mainly differs from international English in phonology, lexicon, and idiom, the same criteria which distinguish, say, New Zealand English from that spoken in Canada’ (1982: 306). If Schmied is nevertheless right in claiming that even spoken African English has more formal features than European varieties then one might predict that pied-piping, the more formal variant, should be more frequent than stranding in almost all contexts.

Finally, the last factor that Schmied mentions is the influence of native-speaker models. He argues that the Kenyan English textbooks for a long time tried to adhere to the exonormative British model, but that there was also influence from other sources (e.g. White South African settlers, moving to Zambia and Kenya; Schmied 1991b: 55). In addition to this, the input model that pupils were exposed to was not the British native speaker. Instead 'a vast majority of English teachers in Africa are … Africans who speak African English themselves, [so] it is not surprising that [the] African dialectal influence is much more dominant than a theoretical British norm, which is still upheld in books but rarely experienced in use in present-day Africa' (1991b: 53). Thus African teachers will perpetuate Africanized versions of English, which would be an indirect way in which L1 transfer could also take place.

Another question that arises from this is the influence of situational factors: as Schmied argues, 'in most parts of Africa nowadays language learning (in classrooms through the conscious study of grammar, etc.) is more important than language acquisition' (1991b: 56). In other words, he emphasizes the classroom setting in which most Africans learn English. He also points out that one reason for the somewhat formal and partly archaic features found in Kenyan English can be attributed to the textbook materials which were used: 'Shakespeare and the Bible have until recently – when they were replaced by modern African classics like Achebe and Meja Mwangi – been most commonly used for teaching the target language' (Schmied 2004a: 925). While this is essentially correct, it should at least be mentioned that from 1970 onwards 'the teaching of English was almost entirely in the hands of Africans, and most textbooks were written and published locally' (Skandera 2003: 13).

In order to asses the validity of Schmied's claim, it was decided to survey the text and grammar books actually used by Kenyan teachers. During my stay in Kenya in 2006, I therefore surveyed the literature available to students of the largest Kenyan university, the University at Nairobi. In addition to modern reference guides like Quirk *et al.* (1985) and Biber *et al.* (1999), the library also stocked a few older grammars from the 1960s and 70s (e.g. Bentley and Sherwood's *English for Modern Africa* (1964), Montgomery's *Effective English* (1971); Palmer's *The Teaching of Oral English* (1974), and Rand's *Constructing Sentences* (1969)). Interestingly, all of these grammars emphasize that preposition-stranding and pied-piping (though they obviously use different terms) are both viable options in modern English, with the former being more informal and the latter being more formal. A case in point is the following passage from Bentley and Sherwood:

(i) This is the boy whom I gave the book to.
(ii) This is the boy to whom I gave the book.
(iii) The trader whom I bought a watch from has been sent to prison.
(iv) The trader from whom I bought a watch has been sent to prison.

As we have learnt, sentences like examples (i) and (iii) with the relative pronoun separated from its preposition are always used in colloquial English. Sentences like examples (ii) and (iv) are used in writing more often than those like examples (i) and (iii). (Bentley and Sherwood 1964: 60–1)

Students training to become teachers thus, at least at university, were given a fairly accurate account of preposition placement in English. Consequently, this should not have led to greater use of the more formal variant, pied-piping, in the Kenyan English speech. (Though note the use of *whom* in all of Bentley and Sherwood's examples, and the lack of a *who* alternative for (i) and (iii).)

Palmer even explicitly draws attention to the fact that stranding is the more natural choice in interrogatives. In his discussion of *wh*-questions he states:

The question here arises as to whether to use the more classical

To whom am I giving then?

or the more casual

Who am I giving them to?

In modern spoken English, WHO with the preposition at the end of the sentence is the current form and the one that the pupils will hear first. For that reason it should be taught here *consistently.* The few remaining natural uses of WHOM (which mainly occur in certain types of relative clauses) should be left until a later stage of progress when differences of usage can be explained. (Palmer 1974: 31–2)

If the reference books available in the library were in fact used for the instruction of prospective teachers, then again this should have led to preposition-stranding being taught to Kenyan pupils as the preferred choice of preposition placement.

Next, it was investigated how preposition placement is dealt with in modern textbooks. The first point to note is that the topic is not addressed during the first eight years of primary education. For the Kenya Certificate of Primary Education (KPCE) exams at the end of year 8, students are only tested on simple relative and interrogative questions (such as *This is the man who gave me the present.* or *Which is the route to Mombasa?*, respectively; sentences taken from *Comprehensive Topical English* 2006: 1).

In order to see how preposition placement is introduced in secondary school textbooks, two widely-used Kenyan textbooks, *Head Start Secondary English* (Bukenya, et *al.* 2003a, 2003; Bukenya, Kioko and Njeng´ere 2004, 2005) and *New Integrated English* (Gathumbi, *et al.* 2003, 2004a, 2004b, 2005), were examined. As it turns out, both textbooks give examples of preposition stranding and pied-piping when discussing interrogative and relative clauses, but do not give any specific information regarding the formality of the two constructions.

In the year 1 book of *Head Start Secondary English* the topic is only indirectly touched upon in the discussion of *wh*-interrogative and *wh*-exclamative sentences. There the expected answers to two questions only have a stranded-preposition alternative in the teacher's book: *What did you cut the tree with?* (Bukenya, *et al.*, 2003a: 83) and *What a mess we are in!* (Bukenya, *et al.* 2003a: 85). The next time the topic comes up is in year 3 in the discussion of the case forms of pronouns. There the following examples are provided under the section 'A pronoun as an object of a preposition' (Bukenya, Kioko, and Njeng´ere 2004: 14):

(3.150) a. To whom do you wish to speak?
b. Did Cody's father tell him whom he wanted to save this letter for?
(examples taken from Bukenya, Kioko and Njeng´ere 2004: 14)

While both preposition placement options are given in (3.150), there is no comment to be found on the rather formal sentence in (3.150a). Note furthermore that in both examples only *whom* and not *who* is offered as the correct pronominal form. This bias for *whom* in object and complement of a preposition function can also be seen in the year 4 book:

(3.151) a. The man, whom you spoke to, is deaf.
b. The ladies, whom I was speaking of, have arrived.
c. This is my cousin, of whom I was speaking.
(examples taken from Bukenya, Kioko and Njeng´ere 2005: 15)

Again, both pied-piped (3.151c) and stranded (3.151a, b) examples are provided, but in all three instances only *whom* and not *who* occurs (note the incorrect use of commas in (3.151a, b)). Thus while the *Head Start Secondary English* books do not seem to prescribe the use of either pied-piping or stranding, they at least exhibit a strong preference for *whom* in both cases. It will therefore have to be examined whether Kenyan English has a stronger preference for *whom* in stranded examples than British English.

Such a preference, however, does not surface in the *New Integrated English* textbooks. In the section on 'Defining and non-defining relative clauses' in the year 2 book, the following set of sentences can be found:

(3.152) a. This is the man to whom you wrote the letter.
b. This is the man you wrote the letter to.
(3.153) The people with whom I live are very pleasant.
(3.154) Labon, who I spoke to on the phone ten minutes ago …
(examples taken from Gathumbi, *et al.* 2004a: 149–50)

Unlike the *Head Start Secondary English* books, the *New Integrated English* textbooks also contain examples with a Ø-relativizer (3.152b) and *who*-relativizer (3.154). Moreover, while all of the above examples with *whom* are pied-piped, the textbook also contains an example with a stranded preposition in the section on 'relative pronouns' (*The driver whom I was with in Garissa has travelled a hundred thousand kilometres*; Gathumbi, *et al.* 2004a: 132).[26]

In both textbooks preposition-stranding and pied-piping are thus presented as possible options, without any further stylistic restrictions. Yet this does not preclude the possibility of individual teachers pointing out the formality of pied-piping to their students. Nevertheless, it will be interesting to see whether the distribution of preposition placement in Kenyan English is as sensitive to the level of formality as British English.

Checking the above list of potential influences on Kenyan English, it turns out that these cover the components situational factors (mostly classroom learning), input (the exonormative British English model and the L2 English of the African teachers and fellow students) as well as learner differences (possible L1-transfer features). Concerning the learner differences not considered so far, i.e. age and motivation, it seems as if these should not have a particular effect on preposition placement.

For most children, English is taught from year 1 and becomes the medium of instruction of all subjects except Swahili from the second part of primary education (Kanyoro 1991: 406; Schmied 2004a: 924; Skandera 2003: 20). This means that by the end of their first four years of primary education, i.e. at the age of 9 or 10, they must have learnt enough of the language to follow their lessons. Kenyans thus learn English from a very early age, but it is difficult to assess how fluent and competent speakers really are in English. Note, however, that in a recent survey by Kioko and Muthwii even 35 per cent of primary school graduates in rural areas, in which competence in English had so far been considered low, reported using English regularly (2004: 37). The social stratum investigated in the present study, i.e. educated Kenyans who have at least completed secondary education, will have had at least twelve years of schooling (eight years of primary and four years of secondary education), in which English was the medium of instruction for at least four years. This group of speakers should thus be fairly fluent in English, and will probably speak it on a daily basis at work, since white-collar jobs in Kenya frequently involve English as the main means of communication (see Kanyoro 1991: 404; Kioko and Muthwii 2004; Schneider 2007: 193). Yet, as indicated above, this fact in itself cannot be said to have any particular effect on preposition placement.

[26] While examples with *that* + P seem to be rare, the student book for year 4 at least contains the following example: *This is the student that we talked about.* (Gathumbi, *et al.* 2005: 23).

The same holds for the motivation factor: since English is strongly associated with upper- and middle-class status and financial success (see Abdulaziz 1991: 393; Kanyoro 1991: 404; Schneider 2007: 193), instrumental (due to the prospect of wealth and status) and integrative (i.e. the will to associate oneself with the upper and middle classes) motivation might be high for the speakers investigated in the present study. Again, no repercussions for the preposition placement preferences of these speakers seem to follow from this observation. Finally, even though individual learner processes (e.g. whether they read many formal English texts like the Bible in their spare time) might have an effect on preposition placement, their precise influence is impossible to assess.

In this chapter various potential factors influencing preposition placement in English were discussed: Clause type (3.1), Type of PP (3.2), Level of formality (3.3) and Type of XP embedded in (3.4). Furthermore it was argued that some of these factors might actually be the result of processing complexity (3.5). If the effect of a factor is indeed processing-based, then it is expected to also surface at least to the same degree in L2 Kenyan English, though this variety might also be affected by other aspects such as a speaker's L1 (3.6). In the next chapter all these claims will be tested against data from the ICE corpora (chapter 4), before specific research questions arising from the multivariate corpus studies are investigated in a series of Magnitude Estimation experiments (chapter 5).

4 Evidence I: Corpus results

4.1 Coding decisions

In the light of the various factors discussed in the preceding chapters, it was decided to code all ICE-GB and ICE-EA data for six contextual factor groups, the first four of which (#1–#4) together with the dependent variable (dv) are presented in Table 4.1. As Table 4.1 illustrates, the main variants of the dependent variable PREPOSITION PLACEMENT were, of course, 'stranded' and 'pied-piped'. In addition to this, an initial survey of the ICE-EA corpus showed that for this variable additional variants had to be included (which are given in parentheses in Table 4.1):

(4.1) a. top managers **with whom** they will be doing business **with**. <ICE-EA:W1B-BK25>[1]

b. Well this is an area **which** for quite some time I've **I've featured on the forefront** violence against women <ICE-EA: S1B037K>

c. Hostels no longer provide conditions **to which** students can study efficiently <ICE-EA: S2B032K>

d. It is a process you have to get initiated **into it** <ICE-EA:S1A026K:B>

Whereas Kenyan English is considered a variety with its own underlying rule system, for the sake of comparison, it is, of course, notable that the tokens in (4.1) deviate from the forms expected from British English: in (4.1a) the preposition *with* occurs both stranded and pied-piped, while in (4.1b) the preposition *of* seems to be missing (cf. *which … I've featured on the forefront* ***of***). In contrast to this, in (4.1c) the preposition *to* surfaces instead of the expected *in* (cf. *conditions* ***in which*** *students can study efficiently*). Finally, occasionally a resumptive pronoun appears behind a preposition which was expected to be stranded ((4.1d); cf. *a process you have to get initiated* ***into***).

[1] Remember that the ICE-GB data was automatically extracted using ICECUP, which provides exact text code information including not only the text category but also the particular line an example is taken from. Since the ICE-EA data was extracted manually, it was only possible to give text category information but not the line number of an example.

Table 4.1 *Dependent variable and factor groups #1–#4*

Factor group		Factors
DV PREPOSITION PLACEMENT		stranded, pied-piped, (doubled), (resumptive), (unexpected), (missing)
#1 CLAUSE TYPES		finite relative, non-finite relative, cleft-relative, main clause question, embedded interrogative, free relative, hollow, passive, comparative, preposed, exclamative
#2 DISPLACED ELEMENT		who, whom, which, Ø, that, what, whose, when, NP, *wh*-ever[a], where, how
#3 TYPE OF XP CONTAINED IN		verb phrase, noun phrase, adjective phrase
#4 TEXT TYPE[b]	[spoken]:	private dialogue, public dialogue, unscripted monologue, scripted monologue, mixed
	[written-as-spoken]:	written-as-spoken
	[written]:	private correspondence, business correspondence, legal presentations, non-professional writing, printed/edited texts

[a] This code included all free relative tokens ending in *-ever*, i.e. *whatever*, *whoever*, *whomever*, *whichever*, *whosoever*, *wherever*, *whenever* and *however*.
[b] Not all of these text types were sampled for both ICE-GB and ICE-EA; see text for details.

As a result, tokens like (4.1a) were classified as 'doubled', like (4.1b) were coded as 'missing' and like (4.1d) were called 'resumptive'. Moreover, tokens like (4.1c) were analysed as 'Unexpected' if the preposition was pied-piped and 'unexpected' if it was stranded.

The first group of independent factors, CLAUSE TYPE, then contains all of the clausal contexts discussed in section 3.1. In addition to these, however, it was also decided to code cleft-relatives separately from ordinary relative clauses due to the specific pragmatic foregrounding function of the former (cf. examples (3.13a–c) repeated here as (4.2a–c)):

(4.2) a. It was John **who** I talked **to**
b. It was John **to whom** I talked
c. It was **to John** that I talked

In order to capture the influence of idiosyncratic pragmatic effects on preposition placement on cleft-relatives, data such as (4.2) were classified as follows: (4.2a) was coded as a stranded cleft-relative and (4.2c) as a pied-piped cleft-relative. In contrast to this, cases such as (4.2b) were treated as pied-piped relative clauses. While this decision might appear somewhat arbitrary, it had the advantage of distinguishing cleft relatives in which a PP was foregrounded (4.2c) from those in which only an NP was highlighted (4.2a). At the same time it made it possible to record the difference between (4.2a) and (4.2b).

The number of factors in group #1 was furthermore increased by subdividing interrogative clauses into 'main clause questions' (*Who did she talk to?*) and 'embedded interrogatives' (*I don't know who she talked to.*). Consequently, it was possible to test whether questions that are embedded in another clause behave differently from main clause questions.

The factors in the next two groups DISPLACED ELEMENT and TYPE OF XP CONTAINED IN then contain no surprises. The group DISPLACED ELEMENT includes the logical complements of the preposition (cf. sections 3.1.1 and 3.1.3), and the group TYPE OF XP CONTAINED IN codes whether the PP in question is embedded in a VP (*Who did she* [*sleep with*]$_{VP}$?), an AdjP (*the things he is* [*capable of*]$_{AP}$) or an NP (*a girl who I couldn't find* [*a present for*]$_{NP}$; cf. section 3.4).

In section 3.3, I claimed that a simple written–spoken dichotomy is insufficient for the evaluation of the effect of formality on preposition-stranding and pied-piping. Therefore, it was decided to use the various ICE TEXT TYPES as factors of the variable formality. As can be seen in Table 4.1, this made it possible to differentiate between formal and informal stylistic levels for both spoken and written English. The only point to remember with this factor group though is that the culture-specific situation in East Africa prevented the ICE-EA team from compiling a corpus which was perfectly matched with the British English component (see section 2.2.1.2). Unscripted monologues, for example, are not part of the Kenyan corpus. On top of that, ICE-EA has two additional text types not found in ICE-GB: 'written-as-spoken' texts (i.e. originally spoken material which had been transcribed by third parties and not the ICE-EA team) and legal presentations (which are formal written manuscripts which are to be read out). Nevertheless, since most text types are identical it was possible to code both corpora for more or less the same factors.

In section 3.2 it was argued that a simple dichotic complement–adjunct classification is inadequate for a detailed syntactic description of the factor group PP TYPE. For this variable it was instead decided to employ the fine-grained classification presented in Table 4.2.

Finally, in the last factor group VARIETY (#6) data were coded as to whether they were taken from British English (ICE-GB) or Kenyan English (Kenyan English subcorpus of ICE-EA).

Thus all corpus tokens were classified according to their variant of the dependent variable PREPOSITION PLACEMENT (dv), as well as the factor groups CLAUSE TYPES (#1), DISPLACED ELEMENT (#2), TYPE OF XP CONTAINED IN (#3), LEVEL OF FORMALITY (#4), PP TYPE (#5) and VARIETY (#6).

Due to the great number of interaction effects discussed in sections 3.1 and 3.1.2.2, only the relative clause data from the corpora were then subjected to an additional multivariate analysis which – with the exception of CLAUSE TYPE – contained all of the factor groups just mentioned. Furthermore these tokens were also analysed for their FINITENESS, RESTRICTIVENESS and COMPLEXITY (Table 4.3).

Table 4.2 *Factor group #5 PP* TYPE *(adapted from Hoffmann 2006: 176)*

Syntactic function of PP		Examples
OBLIGATORY COMPLEMENT	Idiosyncratic stranding Ps	What … for / like
	'V-X-P' idioms	make light of, let go of, get rid of
	Prepositional 'X' (subcategorized P)	sleep with 'have sex with', rely on, capable of
	Subcategorized PP	put something in/on/over
	Obligatory complement	be/live in Spain/on the moon
OPTIONAL COMPLEMENT	Optional complements	work at, talk to, postcards of, a proposal on, worried about
SPACE	Affected location	he sat on the chair, the book on the table
	Movement (goal, source, distance)	he rushed to the church, the paintings from the gallery
	Direction	he ran along the road
	Position/location	he killed the cat in the garden
TIME	Position in time	He died on Saturday, the game on Sunday
	Duration/frequency	He slept for seven hours
PROCESS	Manner	he ate the cake in a disgusting way
	Means/instrument	He killed him with a knife
	Agent	He was killed by John
RESPECT	Accompaniment	He came with Bill
	Respect	For him, something's always missing the article in which she states that… the house with red windows
CONTINGENCY	Cause, reason, purpose, result, condition, concession	as a result of which / due to which
DEGREE	Amplification, diminution …	the extent to which / degree to which

Table 4.3 *Additional factor groups for relative clause analysis*

Factor group	Factors				
FINITENESS	finite, non-finite				
RESTRICTIVENESS	restrictive, non-restrictive				
COMPLEXITY	**2** = <2.5	**a** = 2.5 –<3.0	**3** = 3.0 –<3.5	**b** = 3.5 – <4.0	**4** = 4.0 – <4.5
	c = 4.5 – <5	**5** =5.0 – <5.5	**d** =5.5 – <6.0	**6** = 6.0 – <6.5	**e** =6.5 – <7
	7 = 7.0 – <7.5	**f** = 7.5 – <8.0	**8** = 8.0 – <8.5	**g** =8.5 – <9.0	**9** = >9.0

The first new factor group in Table 4.3, FINITENESS, was introduced because of the categorical effects of non-finite relative clauses (obligatory stranding in Ø-relative clauses and obligatory pied-piping in *wh*-relative clauses; see section 3.1.1). With respect to the token analysis, coding the data for their finiteness is straightforward, since these factors are overtly expressed in a sentence, e.g. by the non-finite marker *to*. In contrast to this, the classification of a relative clause as restrictive or non-restrictive is obviously more difficult.[2] Thus, following Olofsson (1981: 27ff.), in addition to the type of semantic information which a relative clause contributes to the meaning of the antecedent, the following set of criteria were employed to distinguish between the two types of relative clauses.

Non-restrictive relative clauses have weaker semantic ties with their antecedent than their restrictive counterparts. In spoken English, this is often indicated by the insertion of a short pause between antecedent and non-restrictive relative clause (Olofsson 1981: 30). Now, unlike ICE-EA (cf. Hudson-Ettle and Schmied 1999: 13), the spoken ICE-GB data is annotated for short pauses '<,>'. So at least for the ICE-GB data it was possible to use the presence of a pause marker as an indication of a non-restrictive relative clause. In (4.3), for example, the pause, signals that the relative clause is merely a comment. Whereas the presence of such a pause marker turned out to be a reliable indication of non-restrictive clauses, its absence did not allow the identification of a relative clause as restrictive: in (4.4), for example, there is no pause, even though the relative clause only provides additional information about the antecedent:

(4.3) They 've got a **throw-in <,> which they 'll have to settle for** on the far side <ICE-GB:S2A-014 #260:1:A>

(4.4) This is **Humphrey Davy who you may have heard of** in connection with nitrous oxide which he invented <ICE-GB:S2A-027 #1:1:A>

In written English, the pause is mirrored by the orthographic convention of putting a comma between a non-restrictive clause and its antecedent (Huddleston, Pullum and Peterson 2002: 1058). Thus, just as with the pause marker in spoken English, the absence or presence of a comma in the written data was taken as an indicator for restrictive and non-restrictive clauses, respectively. Again, however, the absence of a comma did not necessarily allow the classification of a relative clause as restrictive in either the ICE-GB (4.5) or the ICE-EA data (4.6):

(4.5) a. You will need to show your sight test receipt and your AG 3 to **the person from whom you buy your glasses**. <ICE-GB:W2D-001 #86:1>

b. This novel was followed by ***Shadow of the Condor*** **in which Ronald Malcolm reappears** and was hailed by one critic as 'the most likeable and unlikely CIA agent on record'. <ICE-GB:W2B-005 #59:1>

[2] As discussed in section 3.1.2.2, one should rather see the 'restrictive' vs 'non-restrictive' as an 'obligatory' vs 'non-obligatory' distinction. Due to their wide spread acceptance, it was, however, decided to use the traditional terms.

(4.6) a. It concerns the content of the messages, **the medium through which they are passed** and the mechanisms at work in the passing of such messages. <ICE-EA: W2A016K>
b. Beetles were left for **two weeks after which they were removed by sieving**, leaving larvae to develop. <ICE-EA: W2A024K>

Whereas the relative clauses in (4.5a) and (4.6a) are clearly restrictive, the lack of a comma in (4.5b) and (4.6b) cannot be seen as an indication of restrictiveness: the relative clause might convey important additional information about the novel or the period after the two weeks, but they are clearly optional for the identification of the antecedent's reference.

Furthermore, in accordance with Olofsson's classification (1981: 27ff.), all Ø-relative clauses turned out to be restrictive (cf. (4.7a) for an ICE-GB example and (4.7b) for one from the ICE-EA; also see section 3.1.2.2), as were most *that*-introduced ones. The only exceptions were cases like (4.8b) in which context (cf. 4.8a) has already established the antecedent's reference and the *that*-clause is only functioning as an 'aspect clause', i.e. it is used 'to indicate that a particular aspect of the antecedent is to be thought of' (Olofsson 1981: 29). Interestingly, the data from the ICE-EA contained no examples such as (4.8b).

(4.7) a. I mean wouldn't she have a grown up son and think god he's exactly like **the bloke [Ø] I fell in love with** <ICE-GB:S1A-006 #138:1:B>
b. **The machines [Ø] I talked about earlier** isn't a machine of a person who is able to speak or able to communicate or even able to perceive <ICE-EA:S1B021K:B>

(4.8) a. Actually it's not a small garden …
b. nice nice size garden that she really looks after <ICE-GB:S1A-025 #137:1:B-#141:1:B>

Since it was also possible to check the context of ambiguous examples like (4.8b) in the ICE-GB corpus, the criteria just outlined allowed for a comparatively unproblematic coding for the factor group RESTRICTIVENESS.

As mentioned in section 3.5, it was also decided to test the relative clause data for purely structural complexity effects using Lu's parsing-orientated 'Mean Chunk Number' hypothesis (2002). Since in Lu's approach the Instant Chunk Number (ICN) is divided by the number of words to be integrated, this formula yields continuous variables. Since Goldvarb can only process discrete variables, a continuous variable like the MCN thus has to be arbitrarily divided into discrete categories. Yet, since MCNs are only a heuristic measure of complexity this was not considered to be problematic. In fact, as the statistical analysis showed, it was possible to significantly reduce the categories of this variable (below). On top of that, as Sigley points out, such an arbitrary division of continuous variables is unproblematic for the Goldvarb analysis, 'provided that decisions are made consistently' (1997: 20). In order to guarantee a consistent classification of the various MCNs, standard rounding procedures were employed, which produced the categorical factors as illustrated in Table 4.3. (Example (3.139) *whom I think I had some designs or intentions on*, for example, which was discussed in section 3.5, had an MCN of 4.7. According to the factor divisions given in Table 4.3 this token was coded for the factor 'c'.)

Even though continuous factors are not a problem for the Goldvarb analysis per se, attention must be drawn to the fact that any result involving the factor group COMPLEXITY will nevertheless have to be interpreted carefully. For example, one flaw of the MCN calculation is that complex material has an increased effect, the later it appears in a sentence. Compare e.g. *the knife which*[1] *John*[2] *killed*[3] *the*[4] *man*[4] *with*[5] and *the knife which*[1] *the*[2] *man*[2] *killed*[3] *John*[4] *with*[5], in which both relative clauses contain the same lexical material, and should be expected to be equally complex. However, as the chunk annotation in the square brackets shows, the later the NP *the man* appears in the sentence, the 'heavier' it becomes: thus, the former sentence has an MCN of $(1+2+3+4+4+5)/6 = 3.17$, and the latter $(1+2+2+3+4+5)/6 = 2.83$. Nevertheless, despite these shortcomings, the MCN approach was still considered superior to other measures of complexity since it explicitly predicts that stranding is structurally more complex than pied-piping (see section 3.5).

As the initial run of the Goldvarb program ('no recode') showed, the data extracted from the ICE-GB corpus contained 1,768 relevant tokens, 985 of which were stranded and 783 of which were pied-piped. In contrast to this, the Kenyan data from the ICE-EA had 1,247 tokens, including 808 stranded and 439 pied-piped ones.[3] In addition to this, the Kenyan part of the ICE-EA included 14 doubled-preposition, 22 missing, 18 unexpected (13 of which were stranded, 5 of which were pied-piped) and 7 resumptive tokens. Yet, before comparing mere frequencies it must be kept in mind that the two corpora from which the tokens were extracted differ in size: the ICE-GB corpus consists of 1,060,000 words, while the Kenyan subcorpus of the ICE-EA only has 791,695 words (see section 2.2.1). This is not a problem for the statistical analyses of the data (since Goldvarb and HCFA both successfully correct for such distributional dependence effects; see Sigley 1997: 248–50 and Gries 2008: 247, respectively). For the sake of illustration, however, I will occasionally provide normalized figures of the ICE-EA results to make them comparable to those from the ICE-GB. These normalized figures were obtained by multiplying the ICE-EA figures by a factor of 1.34 (= words in ICE-GB/words in ICE-EA = 1,060,000/791,695).[4]

[3] These figures do not include subject-contained PPs that precede the subject, 17 instances of which can be found in the ICE-GB (*Certain books or scores … **of which details** are given in the leaflet …* <ICE-GB:W2D-006 #100:1>) and 4 in the ICE-EA (*the Mijikenda **of which about** 70% belong to the Giriyama tribal group* <ICE-EA:W2A027K>). As pointed out in section 3.4, stranding is not an option in these cases, but a larger corpus is clearly needed to investigate the factors which lead to preposing such PPs to a pre-subject position.

[4] The normalization calculations were carried out using Excel. The normalized results are always given as full numbers.

Table 4.4 *ICE-GB results for the factor group* CATEGORICAL CLAUSE TYPE

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

Applying this normalization procedure, a one-million word ICE-EA corpus was predicted to contain 1,670 Kenyan tokens, 1,082 of which should be stranded and 588 of which should be pied-piped. Note, however, that in the following, it can be assumed that all figures given are actual frequencies unless explicitly stated otherwise and that only raw frequencies were used for all statistical tests.

Since all the factor groups have been introduced, I will now present the results of the corpus studies. For this I will begin with the categorical clause contexts (4.2), before moving on to the data displaying variation with respect to preposition placement (4.3). Finally, due to their potentially exhibiting additional constraints different to other clause types, I will focus on relative clauses in both varieties (4.4).

4.2 Categorical clause contexts

4.2.1 ICE-GB results

Starting with the ICE-GB data, Table 4.4 shows that 35.2 per cent (=623/1,768) of all preposition placement tokens are clauses which normally require categorical preposition stranding. Surprisingly, as indicated by Table 4.4, the ICE-GB corpus also contains four instances of free relative clauses which seem to contain a pied-piped preposition:

(4.9) a. It's almost like looking into water somehow uhm <,> and as you say it would vary enormously **on what** you put it <ICE-GB:S1B-018 #40:1:B>

b. I feel the most effective means of communicating the key results to respondents is to now enclose a survey summary which pinpoints what UCLi considered to be good practices **from where** we find ourselves at present. <ICE-GB:W1B-029 #10:1>

c. 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>

d. This has tended to obscure **to what extent** Beckett's early writings possess a coherent, though dislocated rhetoric of their own, which Beckett develops, in various ways, sporadically at times, and at others dogmatically, across both some of his critical pieces and his experiments with novel and short story. <ICE-GB:W2A-004 #22:1>

In section 3.1.1 it was pointed out that free relative clauses are normally claimed to be categorical stranding contexts. Yet, all sentences (4.9a–c) are parsed in the ICE-GB as free relative clauses with an initial PP. While the sentence in (4.9a) sounds slightly odd, the examination of its context in the ICE-GB shows that it is a free relative clause whose intended meaning is 'that which [i.e. the surface] you put it [i.e. a painting] on'. Sentence (4.9b), on the other hand, seems grammatical but it could be argued that the preposition *from* can also be parsed as belonging to the matrix clause. In contrast to this (4.9c) contains a *wh-ever* element, which is usually a form associated with free relative clauses. What should be noted about this particular example is the fact that the *wh*-element is embedded as a determiner of an NP (*which-ever country*), which might partly account for the pied-piped preposition (see section 3.4). Interestingly, the last sentence (4.9d) is parsed as an embedded interrogative clause, despite the fact that it corresponds to a relative clause with a pied-piped degree adjunct PP (i.e. *obscure the extent to which* …). The pied-piped preposition in this case can obviously be explained by the categorical effect of this type of PP (see Hoffmann 2005, 2006, and below). Nevertheless the sentences show that free relative clauses might not demand preposition-stranding in all contexts regardless of the type of PP. Due to this observation it was decided to include the free relative clause tokens in the multivariate analysis of the data displaying variation (see section 4.3) in order to see how their effect compares with that of structurally similar clause types such as embedded interrogatives.

Returning to the stranded data it is notable that by far the majority of tokens are finite non-*wh*-relative clauses (348 of 619 tokens, i.e. 56.2%). These consist of a roughly similar number of *that*- (171 instances) and Ø-(177 instances) relative clauses. The next biggest groups of tokens are free relative clauses, which account for 25.5% (158 of 619 tokens), and passives, which make up 15.7% (97 of 619 tokens). Compared to the clause types just mentioned, the number of hollow (14 out of 619 tokens = 2.3%) and comparison clause tokens (3 out of 619 tokens = 0.5%) seems fairly negligible.

Since the above results were obtained from a representative corpus of British English, they allow one to predict to a certain degree with what frequency speakers will encounter stranded prepositions in a particular construction. Finite *that*- and Ø-relative clauses thus appear to be by far the most salient contexts in which speakers will come across obligatorily stranded prepositions. In addition to this, however, another question arising from these data is whether the same types of PPs are licensed in the various clause types. Table 4.5 shows the distribution of PP types across the obligatorily stranding clauses found in the ICE-GB. (Note that all cells for which examples were found in the corpus are shaded light grey in Table 4.5.) As the table shows, the distribution of PP types is fairly complex. Nevertheless, some general patterns can be observed: only optional complement PPs and prepositional verbs occur in all obligatory stranding contexts. In addition to this, V-X-P-idioms are only absent in comparison clauses, which have

Table 4.5 *PP types × obligatorily stranding clauses in the ICE-GB*

PP types	Non-*wh*-RC	Free RC	Passives	Hollow	Comparison	Total
Idiosyncratic P_{strand}	0	23	0	0	0	23
V-X-P idiom	22	8	3	2	0	35
Prepositional V	93	49	61	4	2	209
Subcategorized	3	1	0	0	0	4
Obligatory complement	20	14	0	0	0	34
Optional complement	169	45	31	6	1	252
Affected location	5	1	2	0	0	8
Movement	10	10	0	0	0	20
Direction	2	0	0	0	0	2
Location	5	1	0	0	0	6
Instrument	3	1	0	1	0	5
Accompaniment	14	1	0	1	0	16
Cause	2	3	0	0	0	5
Total	348	157	97	14	3	619

an extremely low overall frequency. Thus, as was to be expected, the great majority of PP types (i.e. (252 + 209 + 35) : 619 tokens = 496 : 619 ≅ 80.1%) surfacing in categorical stranding clause contexts are complement-like in nature. Consequently, it is mainly lexically-stored verb–preposition combinations that are used in categorical stranding constructions.

Turning to the individual clause types, it is not surprising that due to their overall low frequency comparison clauses show the least number of different PP types (only one optional complement (4.10) and two prepositional verbs (4.11a,b)):

(4.10) So I think they're they've put in a pretty pretty good effort **as good as they can hope for** really I think <ICE-GB:S1A-095 #103:1:A>

(4.11) a. I now have a room such **as I always longed for** <ICE-GB:W2B-002 #8:1>
b. this fickle consumer society {which already has more bank and building society branches **than it can cope with**} <ICE-GB:W2C-005 #29:1>

Most of the hollow clauses (6+4 = 10 out of 14 tokens ≅ 71.4%) also contain either optional complement PPs (4.12) or prepositional verbs (4.13):

(4.12) Well the swimming pool's not **worth talking about** <ICE-GB:S1A-021 #142:1:D>
(4.13) Uhm he finds that **very difficult to cope with** <ICE-GB:S1A-076 #45:1:B>

In addition to this, there are also two instances of V-X-P idioms (4.14), one instrument (4.15) and one accompaniment PP (4.16) in the sample:

(4.14) It is a very big oil spill, and that's **quite enough to be getting on with** <ICE-GB:W2B 029 #4:1>
(4.15) Is that still **too difficult to bite with** <ICE-GB:S1A 089 #113:1:A>
(4.16) The Prof is <unclear> <unclear word> </unclear> complete <?> amiableness </?> and his kid, altho' rather obnoxious, is **good fun to play with** <ICE-GB:W1B 011 #112:3>

Furthermore, even though there are considerably more passive tokens in the corpus than hollow clauses, the predicted restrictions on the possible PP types (cf. section 3.1.3) in the former applied in the data: passive sentences only included stranded prepositions which were lexically specified by a verb(al idiom) (e.g. the prepositional verb *deal with* in (4.17), the optional complement *of*-PP in (4.18) or the V-X-P idiom *take care of* in (4.19)) or which headed an affected location PP (e.g. *sat upon* in (4.20)):

(4.17) Maybe his absence is is not properly **dealt with** <ICE-GB:S1B-044 #60:2:B>
(4.18) King 's Canterbury is being **spoken of** very highly at the moment <ICE-GB:S1A-054 #88:1:B>
(4.19) It'll be **taken care of** <ICE-GB:S2A-028 #60:2:A>
(4.20) One of the benches had been **sat upon** <ICE-GB:W2F-005 #97:1>

Now even though free and finite non-*wh*-relative clauses allow for a greater range of PP types, these also mainly exhibit lexically specified PPs: in both free and finite non-*wh*-relative clauses, V-X-P idioms, prepositional verbs, optional complement PPs, obligatory complements and subcategorized PPs make up over two-thirds of all tokens (free relative clauses: (8+49+45+14+1) = 117 of 157 tokens ≅ 74.5%; non-*wh*-relative clauses: (22+93+169+20+3) = 307 out of 348 tokens ≅ 88.2%). Besides this, however, the two constructions also exhibit certain PP types not found in the other three clause types. While an explanation for this might be the overall low token size of hollow and comparative clauses, it is at least noteworthy that speakers should in general encounter a greater range and number of PP types in free and finite non-*wh*-relative clauses. In the categorical ICE-GB data it is only in these clause types that subcategorized PPs (4.21), obligatory complements (4.22), movement (4.23), location (4.24) and cause/reason/purpose PPs (4.25) can be found (note that the free relative clauses are given under (a), while the relative clause tokens are labelled (b)):

(4.21) a. but the environment depends **what situation you're put in** <ICE-GB:S1B-016 #101:1:D>
b. A very uh handy scene of course to fit into a smallish uhm kind of vault **that you can't get uh an overall uh round patterned subject into** <ICE-GB:S2A-060 #58:1:A>

(4.22) a. how can I actually compare what Endsleigh have to offer me in real terms in terms of car insurance to **who I'm with** at the moment <ICE-GB:S1B-074 #3:1:A>
b. it was really the best <?> thing </?> **I've been to** in London <ICE-GB:S1A-098 #206:2:B>

(4.23) a. that's **where the great the greatest enjoyment comes from** <ICE-GB:S1A-001 #40:1:B>
b. Let's take it take it s straight straight in on that and straight on sort of third third stave on the one **I'm coming to** <ICE-GB:S1A-026 #212:1:A>

(4.24) a. He described to me in general detail his business e g **where he operated from** <ICE-GB:S1B-065 #113:1:B>
b. the dance world or the world **that I was working in** <ICE-GB:S1A-001 #35:1:B>

(4.25) a. The great moralist of our time turned out to be Ernest Hemingway when he said 'what's moral is **what you feel good after** and what's immoral is **what you feel bad after**' <ICE-GB:S2B-029 #126:1:A>
b. I suppose that's a source of helium **they <?> need </?> balloons for** now <ICE-GB:S1A-088 #217:1:A>

In addition to this, there are also two types of PPs which only occur either in free relative or finite non-*wh*-relative clauses. Direction PPs (4.26) only surface in non-*wh*-relative clauses:

(4.26) But there's a there's a plain **that the train goes across** <ICE-GB:S1A-011 #86:1:C>

Since these PPs are semantically similar to movement and location PPs, which appear in free relative clauses (cf. (4.23a) and (4.24a) above), it seems safe to assume that this is just an accidental gap in the corpus data.

In contrast to this, the fact that all twenty-three idiosyncratically stranding prepositions (see section 3.2.3.2) only show up in free relative clauses might actually reflect how these are lexically stored: the majority (65%) of these tokens (15 out of 23 instances) are of the form *wh-* X *be like* (4.27a), while the second most frequent combination *wh-* X *look like* only accounts for 17.4% of the data (4 out of 23 instances; (4.27b)).

(4.27) a. I think like an M B A uh in industry and commerce it's as well to have got some feel for what **what life out there is like** <ICE-GB:S1A-035 #111:1:B>
b. Just to show you **what uhm Ionic <,> looks like** …
<ICE-GB:S2A-024 #43:1:A>

One way of capturing the obligatory stranding of *like* would be to assume that the lexicon entry of the preposition contains the information that it must obligatorily be stranded in certain contexts. Yet, data like (4.27) offer an alternative account by attributing the idiosyncratic behaviour of *like* to the fact that it is part of specific, lexically stored collocational idioms such as *wh-* X *be like* (4.27a) or *wh-* X *look like* (4.27b). It will have to be seen whether this analysis receives any support from the data on variable preposition placement.

Table 4.6 *ICE-EA results for the factor group* CATEGORICAL CLAUSE TYPE

Clause type	Stranded	Pied-piped
Finite non-*wh*-relative clauses	249	0
Free relative clauses	130	5
Passive clauses	203	0
Hollow clauses	11	0
Comparison	1	0
Total	594	5

After having investigated preposition placement in the categorical ICE-GB clause types, I will now look at the distribution of these constructions in the Kenyan subcorpus of the ICE-EA. Finally, I will then test whether any detected differences between the two corpora turn out to be statistically significant.

4.2.2 ICE-EA results

As mentioned above, 1,247 preposition placement tokens were extracted from the Kenyan subcorpus of ICE-EA. In Table 4.6 it can be seen that 48 per cent of these (594+5 = 599 out of 1,247 instances) appear in constructions which are considered categorical stranding environments. With respect to their percentage of the overall tokens size, the Kenyan data thus exhibit far more stranded prepositions which occur in categorically stranding contexts than the British data (compare the 35.2 per cent of tokens in the ICE-GB data). Just like the ICE-GB data, however, the Kenyan tokens also include five free relative clauses with a pied-piped preposition:

(4.28) a. They are professing it **in whichever way they want to** <ICE-EA:S1A006K>
b. And therefore it it all depends **from where you are arguing** <ICE-EA:S1A008K>
c. I hope such a thing will not happen and infact, we should strengthen our relationship **in whichever way you'll feel like**. <ICE-EA:W1B-SK47>
d. Depending on how and **by what agency that degradation or change occurs**, the renewability or regeneration of the resources affected cannot always be generated. <ICE-EA:W2A040K>
e. The foregoing statements demonstrate just **to what extent alcoholism is a problem both to the alcoholic and the family**. <ICE-EA:W2B013K>

Two of these examples have a *wh*-ever form ((4.28a) and (4.28c)). In both of these sentences the unit [*in whichever way*] seems to act as a complex free relativizer which links the main and the subordinate clause just as in the German example in (3.59) (section 3.1.4). The free relative clause interpretation of (4.28b), on the other hand, can be inferred by its meaning, which translates into *it all depends on that* [i.e. the position] *which you are arguing*

Table 4.7 *PP types × obligatorily stranding clauses in the ICE-EA*

PP types	Non-*wh*-RC	Free RC	Passives	Hollow	Comparison	Total
Idiosyncratic P_{strand}	0	***0***	0	0	0	0
V-X-P idiom	18	3	24	3	0	48
Pepositional V	99	51	133	4	***0***	287
Subcategorized	***0***	***0***	0	0	0	0
Obligatory complement	19	13	1	0	0	33
Optional complement	86	49	42	2	1	180
Affected location	2	2	3	0	0	7
Movement	6	12	0	0	0	18
Direction	1	0	0	0	0	1
Location	4	***0***	0	1	0	5
Instrument	5	***0***	0	***0***	0	5
Accompaniment	6	***0***	0	1	0	7
Cause	3	***0***	0	0	0	3
Total	249	130	203	11	1	594

from. In (4.38d) the free relative clause introduced by the PP *by what agency* is also selected by *depend on*. Again its paraphrase, *that agency by which* …, supports the free relative clause interpretation of this example. Finally, the degree adjunct PP *to what extent* in (4.28e) corresponds to *the extent to which*, which shows that it is a free relative clause as well. As with the ICE-GB data it was therefore decided to include free relative clauses in the multivariate analysis to compare their effect to other clause types.

As with the British data, the most interesting question for the ICE-EA categorical clausal data is, of course, the types of PPs that can be found in these contexts. Table 4.7 gives an overview of the distribution of PP types across these clause types in the Kenyan component of the ICE project. In order to allow an instantaneous comparison with the British English results cells shaded light grey indicate PP types which were found in the ICE-GB data (cf. Table 4.5 above). Whenever the ICE-EA did not yield an example for such factors, this is highlighted by a bolded and italicized zero (‘***0***’). Alternatively, if a PP type was observed in the ICE-EA which did not surface in the ICE-GB corpus, the relevant cell was shaded dark grey.

It was pointed out above that in the ICE-GB data 80.1 percent of all PP types occurring in categorically stranding clauses were either V-X-P idioms, prepositional verbs or optional complement PPs. As Table 4.7 illustrates, the ratio of such lexically-stored V-P combinations is even higher in the Kenyan data (i.e. 86.7%: (48+287+180) = 515 out of 594 tokens). Whether this slightly higher frequency is significantly higher will be investigated by the statistical analysis below.

Focusing on the individual clause types, the first point to note is that there is only one instance of a comparative clause in the Kenyan subcorpus:

(4.29) Traditionally, members of the community who had **more work than they could cope with** requested assistance from other members of the same community. <ICE-EA:W2A031K>

The V-P sequence *cope with* in (4.29) was classified as an optional complement PP since *cope* can also be used intransitively (cf. e.g. *the reason I think is that uhm modern medicine now enables people to cope* <ICE-GB:S1A-061 #085:1:B>). Nevertheless, *with* is the only preposition that can co-occur with *cope* in its transitive use (cf. **cope on/to/at*), indicating the close association of verb and preposition. Consequently, the fact that the Kenyan comparative tokens do not exhibit a prepositional verb example like the ICE-GB data can be attributed to the low token size (i.e. is only an accidental gap).

As the cell shades show, the second low-frequency phenomenon, hollow clause tokens, pretty much exhibit the same PP types in the Kenyan corpus as in the British one: just like the ICE-GB data, the ICE-EA also includes V-X-P idioms (4.30a), prepositional verbs (4.30b), optional complement PPs (4.30c) and accompaniment PPs (4.30d). In addition to this, while the absence of instrument PP tokens appears to be another accidental gap, the Kenyan data also has one location PP example (4.30e), which is a PP type that was not found in the ICE-GB corpus:

(4.30) a. it's something **worth looking up at** <ICE-EA:S1B031K>
b. Just remember that good men are not **easy to come by**. <ICE-EA:W1B-SK30>
c. At the same time, now that television is no longer novel and offers nothing really **worth talking about** and theatre is mistaken by the authorities as a tool for subversion, those halls might not attract the crowds they used to. <ICE-EA:W2E012K>
d. Once he set his mind on something, he was very hard-working, flexible and **easy to work with**. <ICE-EA:W2B006K>
e. A small bowl or a plate thermoflask to keep water hot a large plastic bowl or *sufuria* with a lid **for storing and sterilizing items in**

Moving on to the passive data, Table 4.7 illustrates that despite the different overall token size, the restrictions on prepositional passives are identical in British and Kenyan English: only lexically stored V-P combinations (such as V-X-P idioms (4.31), prepositional verbs (4.32) or optional complement PPs (4.33)) or affected location PPs (4.34) license prepositional passives. The only new PP type which surfaces in the ICE-EA is the obligatory complement *live in* in (4.35), which is not only an obligatory syntactic element but also has a logical prepositional complement (*two houses*) which can be said to be affected by the verbal action (considering e.g. wear and tear effects):

(4.31) We want the public to be aware that we are very vulnerable and we hate being **taken advantage of** <ICE-EA:S1B034K>
(4.32) The the people who throw out these children if they're found I think they should be **dealt with** <ICE-EA:S1B022K>
(4.33) But the problem is that I think the philosophers very often are not actually **listened to** <ICE-EA:S1B023K>
(4.34) Others will even say: 'WM is **sat on** by his wife these days … <ICE-EA:W2E012K>
(4.35) Why bother building two houses that would only be **lived in** for a short while? <ICE-EA:W2F019K>

Finally, the most common PP types in finite non-*wh*-relative clauses and free relative clauses are, unsurprisingly, lexically specified PPs: V-X-P idioms, prepositional verbs, optional complement PPs and obligatory complements make up 89.2% ((18+99+86+19) = 222 out of 249 instances) of the finite non-*wh*-relative clauses. This proportion is thus comparable to the one obtained from the British English data (88.2%; cf. section 4.2.1). Interestingly, however, in contrast to the ICE-GB data, where there was an almost equal number of *that*- and Ø-relative clauses (171 and 177 tokens, respectively; cf. section 4.2.1), the ICE-EA data exhibits a preference for Ø-relative clauses: the 249 finite non-*wh*-relative clauses consist of 155 Ø- and only 94 *that*-relative clauses. This distributional difference will obviously have to be investigated in the relative-clause-specific studies (see section 4.4).

Comparing the free relative clause data, it turns out that the preference for lexically specified PPs is even greater in the Kenyan English tokens with a percentage of 89.2% ((3+51+49+13) = 116 out of 130 instances) than in the British English ones (74.5%; cf. section 4.2.1). With respect to the remaining PP types, it is noteworthy that subcategorized PPs are not attested in categorically stranding clauses in the ICE-EA, but since these were also rather infrequent in the ICE-GB data (with only three non-*wh*-relative clauses and one free relative clause token) this might just be another accidental gap.

Another distinct difference between the two clause types with respect to the PP types is the following: while location (4.36), instrument (4.37), accompaniment (4.38) and cause/reason/purpose PPs (4.39) occur in non-*wh*-relative clauses in both ICE-EA and ICE-GB (cf. section 4.2.1), in the Kenyan free relative clause data no instances of these PP types surface:

(4.36) no single individual has has a control over the kind of environment **that you are born into** <ICE-EA:S1B031K:D>
(4.37) In Machakos, five people perished on Saturday evening when a Volkswagen beetle **they were travelling in** was swept away in a flooded seasonal river. <ICE-EA:W2C007K>
(4.38) hoping that they can solve Kenyan problems from a different perspective without understanding the people **that they are working with** <ICE-EA: S1BIN5K:A>
(4.39) Today or rather nowadays I actually watch blue movies and do some things **you'll hate me for** if I ever told them to you. <ICE-EA:W1B-SK28>

Again, all of these PP types were infrequent in the ICE-GB free relative clauses (together only accounting for six tokens), so this might also be considered an accidental gap. This seems particularly plausible for location PPs, since the semantically related movement and affected location PPs occur in both non-*wh*-relative (4.40a/4.41a) and free relative clauses (4.40b/4.41b):

(4.40) a. In fact it will also depend on **the community you come from** <ICE-EA:S1A022K:B>

b. **Where I come from** is very hilly <ICE-EA:S2B072K>

(4.41) a. The white cane is supposed to show you that the person who is coming is blind and at the same time it's made in a way such that the sound it produces is able to tell the blind person **the kind of surface he is treading on** <ICE-EA: S1BINT1K:E>

b. The road is a patchy matrix of highs and lows, of huge potholes and loose gravel, of gullies and mocking stretches of smooth tar – bizarre reminders that **what one is actually travelling on** is macadamised road. <ICE-EA:W2B019K>

One point which is, however, notable is the absence of idiosyncratically stranding prepositions from the free relative clause data: while the ICE-GB data contained several idiomatic sequences such as *wh-* X *be like* (4.27a) or *wh-* X *look like* (4.27b), these appear to be less entrenched in the Kenyan free relative clause system.

4.2.3 HCFA: ICE-GB vs ICE-EA

In order to investigate the factors influencing the distribution of stranded prepositions in categorical clausal contexts, the British and Kenyan data were subjected to an HCFA analysis (see section 2.2.2). The variables which were tested were VARIETY (British 'BE' vs Kenyan 'KE' English), PP (the PP types given in Tables 4.5 and 4.7) and CLAUSE (finite non-*wh*-relative clause = 'tzRC', free relative clause = 'free RC', 'passive', 'hollow' and 'comparison' clause). Note that in this and all following HCFAs only the monofactorial results for the variable British versus Kenyan English will be affected by the different sample size of ICE-EA and ICE-GB and the results of these tests will therefore be ignored (which is not really a problem since overall token size is of no importance for the present study). For the interesting interaction of VARIETY with the other variables, HCFA – just like all chi-square tests – corrects for unequal marginals and thus yields valid statistical results (see Gries 2008: 247). In addition, it should be pointed out that the token size of some of the configurations of the HCFAs discussed in this book is relatively low. Therefore in the following only significant effects will be discussed with a coefficient of pronouncedness 'Q' of at least 0.01 (since this is a sample-size-independent measure of effect size (see section 2.2.2); for the full analysis see Appendix A.4.2.3).

Next let us look at the HCFA results for the categorical clausal contexts, starting with the effect of the individual variables. Table 4.8 gives the

Table 4.8 *HCFA for the configurations of the individual variables*

VARIETY	PP	CLAUSE	Freq	Exp	Cont.chi	O-E	P.adj.bin	Q
.	Vprep		496	93.308	1737.918	>	$p < 0.001$	0.360
.	complement$_{optional}$		432	93.308	1229.400	>	$p < 0.001$	0.302
.	direction		3	93.308	87.404	<	$p < 0.001$	0.081
.	subcategorized		4	93.308	85.479	<	$p < 0.001$	0.080
.	cause		8	93.308	77.994	<	$p < 0.001$	0.076
.	instrument		10	93.308	74.379	<	$p < 0.001$	0.074
.	location		11	93.308	72.605	<	$p < 0.001$	0.074
.	affected		15	93.308	65.719	<	$p < 0.001$	0.070
.	accompaniment		23	93.308	52.977	<	$p < 0.001$	0.063
.	idiosyncratic P$_{strand}$		23	93.308	52.977	<	$p < 0.001$	0.063
.	movement		38	93.308	32.783	<	$p < 0.001$	0.049
	complement$_{obligatory}$		67	93.308	7.417	<	$p < 0.05$	0.023
.	.	tzRC	597	242.6	517.722	>	$p < 0.001$	0.365
.	.	passive	300	242.6	13.581	>	$p < 0.001$	0.059
.	.	free RC	287	242.6	8.1260	>	$p < 0.01$	0.046
.	.	comparison	4	242.6	234.666	<	$p < 0.001$	0.246
.	.	hollow	25	242.6	195.176	<	$p < 0.001$	0.224

significant configurations of these analyses. (Note: in this and the following tables, types, i.e. configurations which appear more frequently than expected, are shaded in light grey, while antitypes, i.e. configurations which appear less frequently than expected, are indicated by dark grey cell shading).

The first interesting result of Table 4.8 is that, regardless of the respective variety, it is especially prepositional verbs ('Vprep') and optional PP complements ('complement$_{optional}$') which are significantly associated with categorically stranding clause types. Note also that the coefficient of pronouncedness Q (the sample-size-independent measure of effect size that indicates the amount of variation explained by a particular configuration) of these two factors reveals them as having particularly strong effects (with Q = 0.360 and 0.302, respectively), while the inhibiting effect of the remaining PP types (with Q = 0.081 and below) is already considerably weaker. Finally, within the variable CLAUSE finite non-*wh*-relative clauses, free relative clauses and passive are significantly more frequent than hollow or comparative clauses.

Next, Table 4.9 gives all the significant configurations of two-way interactions. This table indicates that across all clause types, prepositional verbs are significantly more frequent in categorical clause environments in Kenyan

Table 4.9 *HCFA for the configurations of all two-way interactions*

VARIETY	PP	CLAUSE	Freq	Exp	Cont.chi	O-E	P.adj.bin	Q
KE	Vprep	.	287	242.889	8.011	>	$p < 0.05$	0.045
BE	Vprep	.	209	253.111	7.688	<	$p < 0.05$	0.046
KE	.	passive	203	146.909	21.416	>	$p < 0.001$	0.053
BE	.	tzRC	348	304.652	6.168	>	$p < 0.05$	0.048
BE	.	passive	97	153.092	20.552	<	$p < 0.001$	0.053
KE	.	tzRC	249	292.348	6.427	<	$p < 0.05$	0.047
.	Vprep	passive	194	122.671	41.4752	>	$p < 0.001$	0.065
.	idiosyncratic P_{strand}	free RC	23	5.442	56.6506	>	$p < 0.001$	0.015
.	movement	free RC	22	8.991	18.8231	>	$p < 0.05$	0.011
.	Vprep	tzRC	192	244.115	11.1259	<	$p < 0.01$	0.054
.	$\text{complement}_{\text{optional}}$	passive	73	106.843	10.7197	<	$p < 0.05$	0.031
.	$\text{complement}_{\text{obligatory}}$	passive	1	16.571	14.6308	<	$p < 0.001$	0.013

English than in British English. In other words, the HCFA seems to imply that Kenyan speakers prefer stranded prepositions in categorical clauses which in the mental lexicon are strongly associated with the main verb of the sentence. While Kenyan speakers thus heavily employ V-P sequences which are fixed, lexically stored units, British speakers use prepositional verbs significantly less often in these contexts (suggesting that in British English the categorical clause types are much more cognitively entrenched, which allows for a greater range of PP types to surface).

Moreover, Table 4.9 reveals that prepositional passives in Kenyan English are significantly more frequent than expected by chance, while in British English they occur statistically less often than expected. A possible explanation for this finding might be that passives in general are sometimes argued to be favoured in formal styles (cf. Biber 1988: 112, 152). Since Kenyan English is said to employ more formal constructions (cf. section 3.6.3; also Abdulaziz 1991; Schmied 1991b: 53), there might be a general tendency in Kenyan English to use more passives than in British English. This hypothesis will obviously have to be investigated by future studies on East African English.

In contrast to passives, Table 4.9 also shows that finite non-*wh*-relative clauses are statistically more frequent in British English than in Kenyan English. In light of the fact that relative clauses are regarded as the most complex environment for preposition-stranding (see sections 3.1.3 and 3.5), this result can again be explained by the different cognitive status of preposition-stranding in the two varieties: in L1 British English preposition- stranding across all clause types is deeply entrenched in the mental grammars of

Table 4.10 *HCFA for the configurations of all significant three-way interactions*

Variety	PP	clause	Freq	Exp	Cont.chi	O-E	P.adj.bin	Q
KE	Vprep	passive	133	60.071	88.538	>	$p < 0.001$	0.063
BE	complement$_{optional}$	tzRC	169	108.499	33.736	>	$p < 0.001$	0.055
BE	idiosyncratic P$_{strand}$	free RC	23	2.777	147.270	>	$p < 0.001$	0.017
KE	V-X-P idioms	passive	24	10.052	19.353	>	$p < 0.05$	0.012
BE	complement$_{optional}$	passive	31	54.522	10.148	<	$p < 0.05$	0.020

speakers, which facilitates the processing of this construction even in the most complex clausal environments. In L2 Kenyan English, on the other hand, the construction is less entrenched, leading to a lower frequency of preposition-stranding in finite non-*wh*-relative clauses.

Next, Table 4.9 illustrates that the preferred PP type of prepositional passives is, unsurprisingly, prepositional verbs, while optional and obligatory complement appear significantly less often in this clause type. Moreover, free relative clauses appear to favour constructions with movement PPs as well as prepositions which idiosyncratically lead to obligatory stranding ('idiosyncratic P$_{strand}$', e.g. *wh-* X *be like*; cf. the three-way interaction below for a detailed analysis of this effect). In contrast to this, finite non-*wh*-relative clauses disfavour prepositional verbs.

Finally, the significant effects of the three-way interaction configurations are given in Table 4.10. These results can again be considered to reflect the cognitive entrenchment of particular constructions: in Kenyan English prototypical prepositional passives with strongly associated V-P structures (prepositional verbs and V-X-P idioms) are favoured, and, accordingly, cognitively entrenched. In British English, on the other hand, finite non-*wh*-relative clauses are especially associated with optional complement PPs (which in British English seem disfavoured in passives). Finally, the free relative constructions which contain a preposition with idiosyncratic stranding effect can only be said to be entrenched in British English (and, as mentioned above, were not found in the Kenyan data at all; cf. Table 4.7).

This section has shown how the obligatorily stranding clause types in the two varieties resemble each other, for example, in their general preference for lexically associated V-P structures. In addition to this, however, the HCFA also revealed variety-specific differences (e.g. the preference for prepositional passives in Kenyan English or the free relative construction with obligatory stranding in British English). As a next step it will now be necessary to investigate which effects can be found in the variable clause type data from the ICE-GB and ICE-EA corpora.

4.3 Variable clause contexts: ICE-GB vs ICE-EA

As pointed out in section 4.1, the ICE-GB corpus contains 1,768 relevant tokens, 985 of which are stranded and 783 of which are pied-piped, while ICE-EA contains 1,247 tokens (808 stranded and 439 pied-piped ones). Before the Goldvarb analysis could be carried out, however, a large number of tokens with apparent categorical effects ('knockout constraints') had to be either eliminated from the analysis or grouped together ('recoded') with other non-categorical factors from the same factor group (see section 2.2.2). In the following I will first discuss the tokens which were excluded and subjected to HCFAs instead (4.3.1), before turning to the ones that could be recoded (also 4.3.1) and could thus be included in the logistic regression analysis (4.3.2).

4.3.1 Tokens displaying no variation

4.3.1.1 Non-standard ICE-EA tokens

First of all, since in contrast to the ICE-GB component, the Kenyan ICE corpus was read in its entirety, it was possible to retrieve all tokens which exhibit non-standard structures: as it turned out, on top of the above mentioned 1,247 tokens ICE-EA included 14 doubled (cf. (4.42)), 22 missing (cf. (4.43), where *aware of* is expected), 18 unexpected prepositions (13 of which were stranded: e.g. (4.44a), for which *focus on* would be the Standard English collocation; 5 of which were pied-piped: e.g. (4.44b), where Standard English would have *conditions under/in which*) and 7 tokens with a resumptive pronoun (4.45):

(4.42) The thing is **to whom** this drug is going **to** and the use it is going to be put into <ICE-EA:S1B025K>

(4.43) 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:S1B004K>

(4.44) 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:S1B031K>

b. Hostels no longer provide **conditions to which** students can study efficiently <ICE-EA:S2B032K>

(4.45) It is It is a process you have to get initiated into **it** <ICE-EA:S1A026K:B>

One question arising from these examples was whether the variable PPLACE (i.e. whether the preposition was doubled, missing or unexpected, or whether a resumptive pronoun occurred) was significantly associated with specific PP TYPE or CLAUSE TYPE factors. In order to answer this question, these data were subjected to an HCFA (cf. Appendix A.4.3.1.1). This analysis yielded only the significant configurations shown in Table 4.11. There it can be seen that it is mainly prepositional verbs and optional complement PPs which occur in these non-standard structures. This corroborates the claim mentioned in section 3.6.3 that these lexicalized V-P sequences

Table 4.11 *HCFA for the configurations of the individual variables*

PPTYPE	PPLACE	CLAUSE	Freq	Exp	Cont.chi	O-E	P.adj.bin	Q
complement$_{optional}$	.	.	26	4.692	96.758	>	$p < 0.001$	0.378
Vprep	.	.	14	4.692	18.463	>	$p < 0.01$	0.165
.	missing	.	22	12.200	7.872	>	$p < 0.05$	0.201
.	.	finite RC	36	7.625	105.592	>	$p < 0.001$	0.532
.	.	hollow	1	7.625	5.756	<	$p < 0.05$	0.124
.	.	non-finite RC	1	7.625	5.756	<	$p < 0.05$	0.124

often exhibit an idiosyncratic behaviour in Kenyan English (or second-language varieties of English in general). Next, the HCFA shows that it is the absence of expected prepositions ('missing') that is the only significant effect of the PPLACE variable. Finally, non-standard preposition placement phenomena appear most often in finite relative clauses ('finite RC'). This is another unsurprising result in that these clauses are also the most complex ones, which makes them more prone to simplification processes (since missing prepositions are the only significant overall PPLACE effect). The fact that hollow and non-finite clauses ('non-finite RC') seem to occur less frequently in these structures can be accounted for by their low overall frequency.

4.3.1.2 Accidental and systematic gaps I

Returning to the canonical preposition placement structures, the first set of tokens which had to be excluded were the categorical clause types discussed in the previous section. Once the categorically stranding finite-non-*wh* relative clauses, passive, hollow and comparison clause tokens were removed, 523 British (985–348–97–14–3) and 344 Kenyan (808–249–203–11–1) stranded tokens remain. As mentioned above, the set of free relative clauses, despite their alleged categorical effect, were included in the analysis because of the existence of the pied-piped tokens given in (4.9) and (4.28) (section 4.2.1 and 4.2.2). The omission of the non-*wh*-relative clauses was accomplished via the exclusion of all instances of *that* or Ø-displaced elements, which led to a further decrease of the stranded data due to the elimination of 3 *that*-cleft relative clauses and 71 Ø-non-finite relative clauses in the British data (leaving 523–3–71 = 449 stranded ICE-GB tokens) and 47 Ø-non-finite relative clauses in the Kenyan data (leaving 344–47 = 297 stranded ICE-EA tokens).

After the omission of the Ø-non-finite relative clauses, the expected categorical pied-piping effect of the *wh*-non-finite relative clauses surfaced: all 26 British and 14 Kenyan instances of these tokens had a pied-piped preposition, and were therefore also excluded from the subsequent logistic regression analysis (reducing the pied-piped tokens to 783–26 = 757 British and 439–14 = 425 Kenyan ones; for a more detailed discussion of the non-finite relative clauses, see section 4.4.1). It fact, the number of pied-piped Kenyan tokens actually turned out to be 424 at this stage since the following sentence had already been eliminated together with all other Ø-data:

(4.46) The African had also to relate with the natural environment **in he lived** in which he lived in <ICE-EA:S1B027K:B>

This is a performance error, as indicated by the speaker's correction. Without access to the original recording, however, it is difficult to say what exactly the speaker initially produced: either it was the string *in he lived*, which was corrected to a doubled preposition structure *in which he lived in*, or it was *in he lived in* followed by *which he lived in*. While this phenomenon will be looked at in more detail in the Magnitude Estimation experiments (chapter 5), the exclusion of (4.46) could thus obviously be justified on the basis of the speaker's self-correction.

Another set of clause type tokens which had to be eliminated from further analysis were preposed/topicalized clauses. All twelve British and four Kenyan tokens of this clause type exhibited a stranded preposition. For the British data, the lack of pied-piped preposed tokens was the result of the data extraction method since it was not possible to devise an FTF tree within ICECUP that would have allowed the identification of all PPs that have been preposed. Consequently the token set only included those cases in which a preposition was parsed as stranded. In contrast to this, the ICE-EA result was due to the fact that the corpus simply did not contain such structures. Taking a closer look at all stranded preposed tokens, it turns out that the ICE-GB contains six optional complement PPs (4.47a), three V-X-P idioms (4.47b), two prepositional verbs (4.47c; N.B. *going on* here has the meaning 'attend [a party]') and one accompaniment PP (4.47d), while the ICE-EA has two optional complement PPs (4.48a, b), one V-X-P idiom (4.48c) and one adjective–obligatory preposition combination (4.48d):

(4.47) a. Well **Ferndale** I wrote **to** <ICE-GB:S1B-064 #98:1:B>
b. **any idea that the security services were deliberately seeking to bring down the Prime Minister** uh I **give** absolutely **no credence to** <ICE-GB:S1B-040 #88:1:B>
c. So uhm **that** I 'm not **going on** <ICE-GB:S1B-012 #194:1:A>
d. **Harry Beckett** he plays **with** quite a lot <ICE-GB:S1A-058 #265:3:B>

(4.48) a. So **that one I think you don't believe in** <ICE-EA:S1A016K>
b. Striving for equality or whatever is something that we're importing from the west **That one I would disagree with** <ICE-EA:S1A028K>
c. Finally don't worry about the remaining balance. **That will be taken care of** <ICE-EA:W1B-SK04>

d. And **that** one we we we are familiar **with** it[5] a is a word which is formed from the first letters like in our case our university's uh KU Kenyatta University now and we have got many others KISE Kenya Institute of Special Education KIE KIE Kenya Institute of Education <ICE-EA:S2BLEC2K>

While the stranded preposed tokens thus all show a clear preference for lexically associated V/Adj-P sequences, the lack of pied-piped examples prohibited the inclusion of these kinds of clauses in the multivariate analysis (leaving 449–12 = 437 British and 297–4 = 293 Kenyan stranded tokens).

Another clause type with predicted variable preposition placement which had to be excluded was exclamative clauses. In contrast to preposed clauses, full retrieval of all relevant instances of this construction from both corpora was possible (since the entire Kenyan part of the ICE-EA was read and the ICE-GB allowed investigating all stranded preposition tokens and devising a P + *what* FTF in ICE-GB). As it turned out, however, preposition placement in exclamative clauses is an extremely rare phenomenon. In fact, the Kenyan ICE corpus did not even contain a single, relevant exclamative example, while only one token exhibiting a stranded preposition was found in the ICE-GB corpus:

(4.49) **What a mess** she was **in** <ICE-GB:W2F-003 #107:1>

Due to the fact that this singleton instance does not allow for a logistic regression analysis, exclamative clauses were also excluded from the analysis (meaning that only 437–1 = 436 British stranded tokens remained).

Next, it was decided to exclude all tokens with a *wh-ever* displaced element since these data included only one item which appeared with both preposition placement options (the displaced element in question is *whichever*, which in ICE-GB has one stranded (4.50a) and one pied-piped (4.50b) instance, while in ICE-EA only two instances (4.51) with pied-piped prepositions are found).

(4.50) a. You will be surprised how much knowing you are using the best quality bait will do for your confidence and success, **whichever species** you are **after**. <ICE-GB:W2D-017 #38:1>

b. 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>

(4.51) a. They are professing it **in whichever way they want to** <ICE-EA:S1A006K>

b. I hope such a thing will not happen and infact, we should strengthen our relationship **in whichever way you'll feel like**. <ICE-EA:W1B-SK47>

All other *wh-ever* displaced elements exclusively co-occur with stranded prepositions (in the ICE-GB comprising five *whatever* (4.52a), two *wherever*

[5] Note that this example was analysed as [***that** one we we we are familiar **with***] [*it a is a word* …]. Without access to the original recordings it is, however, also possible that the pronoun *it* functions as a resumptive pronoun of *with*.

(4.52b) and one *however* (4.52c) tokens, and in the ICE-EA consisting of five *whatever* (4.53) and one *wherever* tokens (4.54)):

(4.52) a. **whatever subject** you're interested **in** you could then follow that through <ICE-GB:S1B-025 #160:1:C>
b. We have to attack it and say to everybody in this country whatever they feel **wherever** they come **from** if they share the beliefs that we have in the Conservative Party then they we don't only re uh uh hope that they will vote for us <ICE-GB:S1B-043 #155:1:B>
c. so **however** we **many** we decide **on** I'm sure we'd want some postgraduates as well as undergraduates <ICE-GB:S1B-075 #54:1:A>

(4.53) **Whatever you want to talk about** <ICE-EA:S1A013K>

(4.54) Exactly when Mutiso came from[6] **wherever he claimed to have come from** was also another mystery… <ICE-EA:W2F010K>

4.3.1.3 Idiosyncratic effects: Obligatorily stranded prepositions and the effect of to be

The exclusion of these *wh-ever* items leaves 427 stranded (= 436–1–5–2–1) and 756 pied-piped (= 757–1) British and 287 stranded (= 293–6) and 422 pied-piped (= 424–2) Kenyan tokens. This reduced data set then revealed several knockout constraints in the factor group PP TYPES, as follows.

As expected, all instances of prepositions with an alleged idiosyncratic stranding effect did in fact only exhibit stranded prepositions: sixty-one such tokens were found in ICE-GB. In addition to the twenty-three free relative tokens already mentioned in section 4.2.1, these idiosyncratically stranded prepositions from the British corpus occurred in questions (twenty tokens; e.g. (4.55a)), embedded interrogatives (seventeen tokens; e.g. (4.55b)) and relative clauses (one token: (4.55c)):

(4.55) a. Yeah you still imagine **what it was like** <ICE-GB:S1A-016 #300:1:E>
b. **What's it like** <ICE-GB:S1A-019 #53:1:B>
c. To meet **someone you look like** and whose blood and characteristics you have … <ICE-GB:W1B-003 #122:2>

In contrast to this, only five such prepositions were found in the Kenyan subcorpus of ICE-EA:

(4.56) a. **How's your week-end like** or how was it <ICE-EA:S1BINT3K>
b. Tell us uh briefly what then **what has been the response like** for the last few days you've staged The Lion and the Jewel <ICE-EA:S1BINT3K>
c. and you wonder why didn't he go to the small kitchen or to a store and look around and go to the supermarket himself instead of asking **what is it like** you don't have here <ICE-EA:S1A030K>
d. Sometimes when some of us think about the future we wonder **how the country shall be like** when these confused people will be elder <ICE-EA:S2B040K>
e. **What** did she want to remind Kanaya of these other possibilities **for**? <ICE-EA:W2F008K>

[6] Note that the preposition *from* which precedes wherever in (4.54) is being licensed by the main verb *come*. This was therefore not considered an instance of a doubled-preposition structure.

As (4.56) shows, four out of the five instances with idiosyncratically stranding prepositions are of the *wh- be like* type (though note that (4.56a) has a *how* displaced element, which would be unidiomatic in British English).

Due to their categorical effect, tokens with such prepositions also had to be excluded from the Goldvarb analysis, which meant that only 367 stranded ICE-GB tokens (= 427–60; the relative clause example (4.55c) had already been excluded earlier along with all other Ø-tokens) and 282 stranded ICE-EA (= 287–5) tokens remained. Nevertheless, while these data were therefore omitted from further logistic regression analyses, a closer inspection of these tokens indicated that they needed to be subjected to further analysis: over 85.2 per cent (fifty-two out of sixty-one instances) of the British tokens and 80 per cent (four out of five) of the Kenyan tokens were of the type *wh-* X *be like*. Thus it seems as if this string is indeed a lexically stored collocational idiom. In order to test this hypothesis, it became necessary to compare statistically the distribution of this construction with all other instances of preposition placement with *be*. As it turned out, all fifty-three *be*-tokens in the obligatory complement PP group in the ICE-GB corpus also only occurred with stranded prepositions. The ICE-EA data, on the other hand, contained forty-nine *be* tokens, forty-six of which were stranded, but three of which exhibited pied-piping:

(4.57) a. Speaker Jonathan Ng'eno noticed the tricky situation **in which Mr Onyancha was** and decided [t]o end his agony by ordering him to carry out further <ICE-EA:W2E014K>
b. 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:S1B067K>
c. **For what purpose is it for** <ICE-EA:S1B041K>

In (4.57a) the preposition is pied-piped, despite the fact that in British English only stranding would be considered acceptable (cf. sections 3.2.3.2, 4.3.1.1). This stranding constraint of *be* seems to be responsible for the doubled-preposition structures in (4.57b, c): the speakers had already produced a clause-initial pied-piped preposition (*in which* and *for what purpose*, respectively) when *be* was realized, whose lexical constraints require a following stranded preposition, which is satisfied by the insertion of a doubled preposition.

Taking a closer look at all *be*-tokens shows that the idiosyncratic stranding effect of this verb (see section 3.2.3.2) can be observed across all clause types: with the exception of the three tokens in (4.57), all free relative clauses (fourteen ICE-GB instances, e.g. (4.58a); thirteen ICE-EA instances, e,g. (4.58b)), questions (eleven ICE-GB instances, e.g. (4.59a); eight ICE-EA instances, e.g. (4.59b)), embedded interrogatives (twelve ICE-GB instances, e.g. (4.60a); thirteen ICE-EA instances, e.g. (4.60b)), relative clauses (fifteen ICE-GB instances, e.g. (4.61a); twelve ICE-EA instances, e.g. (4.61b), including one cleft-relative: (4.61c)) and exclamatives (one ICE-GB instance: (4.49), repeated here as (4.62)) only exhibit stranded prepositions:

(4.58) a. That's not **what linguistics is about** <ICE-GB:S1A-038 #313:1:C>
b. that philosophy uh makes people riot or things like that which is not really **what philosophy is all about** <ICE-EA:S1B023K>

(4.59) a. **Who is it by** <ICE-GB:S1A-043 #68:1:A>
b. **Which school were you in** <ICE-EA:S1A006K>

(4.60) a. And Dad Dad walked in half an hour later saying I've got these two lines of poetry going through my mind and I don't know **where it's from** <ICE-GB:S1A-032 #169:1:B>
b. Well to begin with I would like to ask Dr Gikenye to tell us something in brief **what ma medicine is all about** <ICE-EA:S1B021K>

(4.61) a. the guy **I was with** <ICE-GB:S1A-052 #89:2:B>
b. A wife who is a widow uh may decide whether or not even according to the Luo traditional customs whether or not she will be inherited quote unquote And if she decides she will be inherited she decides among the brothers of **the clan she is in** <ICE-EA:S1A009K:D>
c. **It's not durability I'm after**. I don't mind buying new shoes every summer if they're fashionable <ICE-EA:W2F009K>

(4.62) **What a mess** she was **in** <ICE-GB:W2F-003 #107:1>

The obligatory stranding effect of *be* might, of course, be partly due to its semantic weakness (cf. Trotta 2000: 62). Yet, since *be* can occur at the end of clauses such as *I don't know who he is*, this effect cannot simply be attributed to semantic weakness. Instead, a lexically stored constraint must be assumed that requires *be* to co-occur with stranded prepositions only. Because of this overall constraint, it is very well possible that the preposition *like* has such a strong idiosyncratic stranding effect due to its frequent collocation with *be* (see above).

As with the idiosyncratic preposition tokens, the *be*-data had to be eliminated from the Goldvarb analysis, which reduced the number of stranded ICE-GB tokens to 330 (= 367–14–11–12)[7]. For the stranded ICE-EA tokens this led to a reduced data set of 246 (= 282–36: ten non-*wh*-relatives with *be* had already been excluded earlier; the exclusion of the one pied-piped example gave 422–1 = 421 tokens). With respect to the lexical stranding effect of *be*, however, an interesting question arose from the above observations: are there any specific *be* + $\mathrm{P}_{\text{stranded}}$ sequences which can be shown to be statistically associated with a particular clause type?

In answer to this question all 105 ICE-GB *be*-tokens (52 with an idiosyncratically stranding preposition, 53 with a PP functioning as obligatory complement) and all 50 ICE-EA *be*-tokens (4 with an idiosyncratically stranding preposition, 46 with a PP functioning as obligatory complement) were subjected to a HCFA. In this analysis, the factors VARIETY (British vs Kenyan English), CLAUSE (relative clause, free relative, etc.) and P (the particular preposition) were tested. As before, Table 4.12 first of all gives the significant configurations of the mono factorial HCFA (the full analysis can be found inAppendix A.4.3.1.3).

[7] The fifteen relative clauses (thirteen finite non-*wh*- and two non-finite Ø-relatives) as well as the exclamative example (4.49) had already been omitted from the overall analysis earlier (see above).

Table 4.12 *HCFA result for be-tokens*

VARIETY	CLAUSE	P	Freq	Exp	Cont.chisq	O-E	P.adj.bin	Q
	interrogative$_{embedded}$		44	26	12.462	>	$p < 0.01$	0.138
	question$_{main\ clause}$		43	26	11.115	>	$p < 0.01$	0.131
	free RC		41	26	8.654	>	$p < 0.01$	0.115
	cleft		1	26	24.039	<	$p < 0.001$	0.192
	exclamative		1	26	24.039	<	$p < 0.001$	0.192
		like	56	12	161.333	>	$p < 0.001$	0.306
		about	43	12	80.083	>	$p < 0.001$	0.215
		in	29	12	24.083	>	$p < 0.001$	0.118
		after	1	12	10.083	<	$p < 0.001$	0.076
		at	1	12	10.083	<	$p < 0.001$	0.076
		by	1	12	10.083	<	$p < 0.001$	0.076
		of	1	12	10.083	<	$p < 0.001$	0.076
		upto	1	12	10.083	<	$p < 0.001$	0.076
		to	2	12	8.333	<	$p < 0.01$	0.069
		from	3	12	6.750	<	$p < 0.05$	0.062
		on	3	12	6.750	<	$p < 0.05$	0.062
KE		like	4	18.308	11.182	<	$p < 0.001$	0.104
	RC	in	21	4.833	54.075	>	$p < 0.001$	0.107
	RC	like	0	9.333	9.333	<	$p < 0.01$	0.064
	free RC	in	0	7.622	7.622	<	$p < 0.05$	0.051
BE	RC	in	13	3.253	29.202	>	$p < 0.01$	0.064
KE	RC	in	8	1.580	26.0839	>	$p < 0.05$	0.042

The HCFA thus shows that it is especially in embedded interrogatives, free relatives and questions that these sequences appear. Furthermore, the prepositions most closely associated with *be* are *like*, *about* and *in*. In addition to that, the significant two- and three-way factor configurations also contain several interactions: as the VARIETY × P interaction illustrates, *be like* idioms turn out to be disfavoured in Kenyan English (i.e. they are not as deeply entrenched as in British English). Moreover, *be in* is especially entrenched in relative clauses, a conclusion that receives support from the three-way interaction results: since both varieties exhibit similar behaviour with respect to this structure, it might be that this is the result of a functional/pragmatic constraint: *be*-relative clauses might be particularly frequent due to the fact that people generally speak more about the places they have been to or where something of interest is located.

Table 4.13 *Selected PP types with expected and actual knockout effect*

	ICE-GB			ICE-EA		
PP type	stranded	pied-piped	total	stranded	pied-piped	total
Frequency	7	27	34	0	14	14
Manner	2	86	88	0	44	44
Respect	0	134	134	0	78	78
Degree	0	34	34	0	19	19
Direction	0	4	4	0	3	3
Agent	0	0	0	0	3	3
Subcategorized PPs	1	7	8	0	3	3
Instrument	10	42	52	0	49	49
Cause/reason/result	6	23	29	0	24	24
Total	26	357	383	0	237	237

4.3.1.4 Categorical PP types: Accidental and systematic gaps II

Returning to the tokens remaining for the main analysis (ICE-GB: 330 stranded and 756 pied-piped; ICE-EA: 246 stranded and 422 pied-piped), in light of the results from an earlier study on relative clause data (Hoffmann 2005, 2006) several additional PP types were expected to exhibit obligatory pied-piping, namely respect (*you may only use this* ***under certain conditions***), manner (*she killed him* ***in a very brutal way***), frequency/duration (*they slept* ***for three hours***) and degree adjuncts (*they accepted the pay cuts* ***to a certain extent***). The reason for the categorical behaviour of these PP adjuncts was argued to be that stranding a preposition with these leads to an uninterpretable V-P_{stranded} complex. For manner and degree adjuncts such an explanation receives support from the fact that these do not add thematic participants to a predicate but compare events 'to other possible events of V-ing' (Ernst 2002: 59). Furthermore, frequency/duration adjuncts and respect adjuncts do not contribute thematic participants to a predicate either. Instead, the former have scope over the temporal information of an entire clause, while the latter identify the 'relevant point of reference in respect of which the clause concerned derives its truth value' (Quirk *et al.* 1985: 483).

As a look at the distribution of the PP types in the two corpora shows, however, the predictions concerning the obligatory pied-piping effect of these PP adjuncts are only partly confirmed. Table 4.13 gives the results for those PP types with an expected and those with an actual knockout effect in the corpus; it shows that the predicted effect can be found in the ICE-EA data: all respect (4.63), manner (4.64), degree (4.65) and frequency/duration (4.66) adjunct PPs only co-occur with pied-piped prepositions.

(4.63) So who are those Who are the three classes Philosophers **amongst who** we get the philosopher king okay <ICE-EA:S2BLEC1K>

(4.64) This approach directs attention to how the members of a given society are related to their environment – particularly to how they exploit it and to the network of social relations involved – in an effort to determine **in what ways** these relations illumine other aspects of their social system. <ICE-EA:W1A007K>
(4.65) **To what extent** have workers in the private or public sector been kicked out of their jobs for reasons other than incompetence or gross insubordination? <ICE-EA:W2E013K>
(4.66) depends on: – quantity of the drug consumed per occasion – **frequency with which** that quantity is consumed <ICE-EA:W2A023K>

In contrast to this, in the ICE-GB data only respect (4.67) and degree (4.68) PPs exhibit the expected categorical pied-piping effect:

(4.67) **in what respect** do you love us <ICE-GB:S2A-036 #17:1:A>
(4.68) Well if in humans there is a genetically encoded part of language uh you'll have to find out ba **to what extent** uh language is genetically encoded <ICE-GB:S1B-003 #35:1:B>

Moreover, both corpora contain categorical cells that had not been expected: direction PPs e.g. (4.69) and (4.70) do add a thematic entity (in the sense of Ernst 2002) in that they specify the path along which the action denoted by motion verbs proceeds, so that the lack of stranded tokens with these PPs appears to be an accidental gap in the data.

(4.69) From his window the young boy would have looked across the green fields to Camden Town, and to the Hampstead Road **along which** the old stage coaches still travelled. <ICE-GB:W2B-006 #62:1>
(4.70) Now what I was wondering Doctor is the role of culture in development because you find that the path **towards which** we are moving is a path of development as defined by the current you know the trends … <ICE-EA:S1B043K>

This raises the question, however, how one can distinguish accidental gaps (which should remain in the data set by e.g. recoding them with other suitable factors) from systematic ones (which should be removed from the statistical variation analysis). As I have pointed out elsewhere (Hoffmann 2006, 2007b), in addition to corroborating experimental data (cf. chapter 5), intra-corpus evidence can also sometimes prove helpful: take, for example, the apparent categorical effect of direction PPs. As a closer look at the tokens that had been excluded earlier showed, these PPs occur in categorical stranding environments in the corpus (cf. (4.71) and (4.72)):

(4.71) But there's a there's **a plain that the train goes across** <ICE-GB:S1A-011 #86:1:C>

(4.72) The difficult task of keeping to your seat or negotiating the car over the yawning gaps means that many people miss the enchanting beauty of **the area they are travelling through**. <ICE-EA:W2B019K>

Examples like (4.71) and (4.72) had already been excluded from the data set before the compilation of Table 4.13 due to the presence of non-*wh*-relativizers, which categorically induce stranding. Their grammaticality, however, indicates that preposition-stranding with direction PPs in the remaining data set is not generally impossible but merely an accidental gap.[8] Consequently, direction PPs were not excluded from the subsequent analyses but recoded into location adjunct PPs due to the similar semantic function of the two PP types.

A similar argument can be made for the effect of instrument and cause/reason/result PPs in the Kenyan data of Table 4.13: not only do these PPs occur with a stranded preposition in the ICE-GB corpus, but even ICE-EA has non-*wh*-relative clauses in which these PPs strand their preposition (cf. the instrument PP in (4.73) and the cause/reason/result PP in (4.74)). These PP types were therefore kept in the Goldvarb analysis (and, following several statistical tests to identify the optimal set of recodings, grouped together with a number of other adjuncts; for details see section 4.3.2).

(4.73) One of the two oxen **she had <+_been> ploughing her *shamba* with** had to go. <ICE-EA:W2C025K>

(4.74) Today or rather nowadays I actually watch blue movies and do some **things you'll hate me for** if I ever told them to you. <ICE-EA:W1B-SK28>

In contrast to this, the ICE-EA corpus contained no stranded preposition tokens with subcategorized PPs. However, there was still reason to consider this an accidental gap: first of all, this PP type has an overall low frequency (only three tokens in ICE-EA), which might explain the accidental nature of this phenomenon. On top of that, the only slightly bigger British data set (eight tokens) already contains a stranded example:

(4.75) but the environment depends **what situation you're put in** <ICE-GB:S1B-016 #101:1:D>

Example (4.75) is a free relative clause, which will definitely have had a strong stranding effect in this example. Nevertheless, it also shows that subcategorized PP tokens can be stranded and that the Kenyan data in Table 4.13 probably reveal only an accidental gap.

Table 4.13, however, also includes another PP group which I decided to exclude from the Goldvarb analysis, namely the following three agentive ICE-EA tokens (4.76a–c):

(4.76) a. Now <name/> can you tell us about songs when where and **by whom they are performed** <ICE-EA:S2B075K>

b. Can the Assistant Minister tell us when he is going to table this bank statement, so that we may know when the money was banked in the account, how much was withdrawn, **by whom it was withdrawn** and the signatories. <ICE-EA:S1BHN14K>

[8] This interpretation is supported by the fact that direction PPs in general are fairly infrequent in both corpora (cf. the four British and three Kenyan tokens in Table 4.13). Note, however, that there is still the possibility of an interaction effect, i.e. that preposition-stranding with direction PPs is only entrenched in e.g. non-*wh*-relative clauses. While this may at first glance seem somewhat unlikely, it might nevertheless warrant further experimental investigation.

c. Where it is recorded that there was interpretation at the request of the Appellant, and into what language there was interpretation and **by whom it was done**, there is no problem. <ICE-EA:W1C001K>

The reason for excluding these tokens had to do with the comparability of the corpus studies. The ICE-GB corpus has no agentive tokens at all, while ICE-EA only contains the above three tokens. I therefore found it impossible to decide whether the categorical effect in Table 4.13 should be seen as an accidental or systematic gap. I suspect that *by*-agents should have a strong pied-piping preference since they are not associated with particular verbs, but constitute an optional phrase linked to the abstract passive construction. Yet this is a claim that will definitely require future empirical analysis.

Returning to the PP types which were hypothesized to systematically trigger categorical pied-piping, the fact that the ICE-GB corpus contained stranded manner and frequency adjunct PPs was somewhat unexpected. However, once the stranded instances of these PP types were examined in more detail, these turned out to be another case of lexically idiosyncratic effects. Beginning with the stranded manner tokens, (4.77) gives both instances that surface in the ICE-GB corpus:

(4.77) a. But **which order do they come in** <ICE-GB:S1B-015 #70:1:C>
b. And she said uh it's very easy to remember **which order they come in** if you actually just think simply think of their initials <ICE-GB:S2A-024 #35:1:A>

While *in which order* in both examples clearly specifies the manner (i.e. how) an event takes place, they should not have been classified as optional adjunct PPs. Instead, the VP *to come in an order* is an idiomatic collocation which can be argued to be a complex lexical item. Consequently, stranding *in* in (4.77) is unproblematic since the processor can predict the preposition once the string *order … come* has been parsed. The remaining eighty-six manner PPs in the corpus for which no such lexicalized explanation is available show the expected categorical pied-piping constraint (as in (4.78)):

(4.78) I could talk quite a long time about my father on a question like this and **in what kind of way** do you want him described as a as a person <ICE-GB:S1A-076 #32:1:B>

Concerning frequency PPs, the data also suggest a lexicon-based explanation since all seven British stranded examples occur with the same displaced element (*how*) and the same preposition (*for*; cf. the six question (4.79) and the single embedded interrogative (4.80) examples below):

(4.79) a. **How long** did you do English **for** <ICE-GB:S1A-006 #1:1:A>
b. Uhm **how long** do they go on **for** <ICE-GB:S1A-011 #46:1:B>
c. **How much longer** do we seriously have to talk **for** <ICE-GB:S1A-038 #268:1:A>
d. Well so **how long** did you live in Portugal **for** <ICE-GB:S1A-041 #197:1:B>

e. And **how long** are you there **for** <ICE-GB:S1A-097 #68:1:B>
f. And uh **how long** did that go on **for** <ICE-GB:S1B-066 #81:1:A>

(4.80) so that you know **how long** you're going to be sitting there **for** before you find out what happens to Oedipus<ICE-GB:S1B-019 #22:1:A>

This appears to imply that *how long ... for* is a complex, discontinuous lexical item in British English, which again allows the parser to process the stranded preposition despite the fact that frequency adjunct PPs per se are difficult to interpret.

Note, however, that there is nevertheless a distinct difference between the stranded frequency and manner adjunct PPs: while the manner PP headed by *in* in (4.77) is obligatorily selected by a particular verb (*come*), the discontinuous frequency PP is more adjunct-like in that it co-occurs with a wider range of predicates (*do* in (4.79a), *go on* in (4.79b, f), *talk* in (4.79c), *live* in (4.79d), *be* in (4.79e) and *sit* in (4.80)). This adjunct-like status also accounts for the fact that the ICE-GB corpus contains a pied-piped version of this idiomatic syntagm:

(4.81) It was impossible to say **for how long** John Dickens might be incarcerated <ICE-GB:W2B-006 #16:1>

Even more interestingly, the fourteen Kenyan frequency tokens also included the following four *how long*-questions:

(4.82) a. And uh **for how long** have these child rights clubs been in existence <ICE-EA:S1B049K>
b. But **for how long** will Ethiopia remain a receiver of something as basic as food because of a seemingly endless battle? <ICE-EA:W2E008K>
c. So, the teacher starts wondering, 'What is wrong with me? Am I in the wrong profession? Am I in the suffering profession such that I keep on suffering, and **for how long** shall I suffer.' <ICE-EA:S1B055HK>
d. And the dark nights? **For how long** were people to grope their way along unlit streets? <ICE-EA:W2B020K>

Thus, while in the British data *how long*-tokens predominantly occur in the discontinuous *how long ... for* form (cf. (4.79) and (4.80)), in the Kenyan data only the pied-piped structure surfaces (4.82a–d). The reason for this might be a combination of processing effort and lexicalization: frequency adjunct PPs with a stranded preposition are difficult to parse since they have scope over the temporal information of the entire clause. Yet since *how long ... for* is entrenched as a complex lexical item in British English, the processing effort is significantly reduced. In Kenyan English, on the other hand, the discontinuous structure appears less entrenched, which means that it would have to be parsed on-line without any support from the lexicon. Consequently, pied-piping is preferred in this variety.

As a result of the above discussion, it was decided to keep all frequency PP tokens for the subsequent analysis (due to the adjunct-like character

Table 4.14 *HCFA result for individual variables in obligatory PP types*

VARIETY	PP	P_{place}	CLAUSE	Freq	Exp	Cont.chisq	O-E	P.adj.bin	Q
.	respect	.	.	212	110.75	92.565	>	$p < 0.001$	0.305
.	frequency	.	.	48	110.75	35.554	<	$p < 0.001$	0.189
.	degree	.	.	53	110.75	30.113	<	$p < 0.001$	0.174
.	.	piped	.	436	221.5	207.721	>	$p < 0.001$	0.968
.	.	stranded	.	7	221.5	207.721	<	$p < 0.001$	0.968
.	.	.	RC	391	88.6	1032.119	>	$p < 0.001$	0.853
.	.	.	free RC	1	88.6	86.611	<	$p < 0.001$	0.247
.	.	.	$\text{interrogative}_{\text{embedded}}$	9	88.6	71.514	<	$p < 0.001$	0.225
.	.	.	cleft	18	88.6	56.257	<	$p < 0.001$	0.199
.	.	.	$\text{question}_{\text{main clause}}$	24	88.6	47.101	<	$p < 0.001$	0.182

of the stranded tokens as well as the existence of pied-piped alternatives for these). In contrast to this, all respect, degree and manner PP adjuncts were eliminated due to their inherent categorical pied-piping effect (with the stranded manner PPs in (4.77) being treated as lexical idiosyncrasies). This final exclusion of factors reduced the data set for the multivariate analysis to

- 830 British tokens: 328 stranded (330 – 2 idiosyncratic manner PPs) and 502 pied-piped (756 – 86 manner – 134 respect – 34 degree PPs).
- 523 Kenyan tokens: 246 stranded (see above) and 277 pied-piped (421 – 44 manner – 78 respect – 19 degree – 3 agent PP)

4.3.1.5 Obligatorily pied-piping PP types

Given the categorical effect of degree, manner, respect and frequency PPs, it was decided to subject these data to a HCFA to investigate whether these also display clause-type specific idiosyncratic constraints. The variables included in this analysis were VARIETY, PP TYPE, PREPOSITION PLACEMENT (pied-piped vs stranded; this became necessary since the stranded frequency tokens were also included) and CLAUSE TYPE. Table 4.14 summarizes all configurations of the individual variables which emerged as significant in this analysis (for the full HCFA, see Appendix A.4.3.1.5). From there it can be deduced that respect adjunct PPs are by far the statistically most frequent of the PP types, while degree and frequency PPs occur less often than expected. Moreover, the results for preposition placement ('Pplace') corroborate the strong pied-piping effect of these PPs (note the extremely high coefficient of pronouncedness 'Q' of 0.968). Finally, all PPs occur by far more frequently in relative clauses than all other clausal contexts.

Table 4.15 *HCFA result for all significant interactions in obligatory PP types*

VARIETY	PP	P_{place}	CLAUSE	Freq	Exp	Cont.chisq	O-E	P.adj.bin	Q
KE	.	.	cleft	0	6.298	6.298	<	$p < 0.05$	0.014
.	frequency	stranded	.	7	0.759	51.360	>	$p < 0.001$	0.014
.	frequency	.	$\text{question}_{\text{main clause}}$	11	2.601	27.130	>	$p < 0.01$	0.019
.	.	stranded	$\text{question}_{\text{main clause}}$	6	0.379	83.316	>	$p < 0.001$	0.013
.	.	stranded	RC	0	6.178	6.178	<	$p < 0.05$	0.014
BE	frequency	stranded	.	7	0.493	85.864	>	$p < 0.001$	0.015
BE	.	stranded	$\text{question}_{\text{main clause}}$	6	0.247	134.291	>	$p < 0.001$	0.013
KE	.	piped	cleft	0	6.199	6.199	<	$p < 0.05$	0.014
.	frequency	stranded	$\text{question}_{\text{main clause}}$	6	0.041	863.954	>	$p < 0.001$	0.013
BE	frequency	stranded	$\text{question}_{\text{main clause}}$	6	0.027	1336.341	>	$p < 0.001$	0.013

Next, Table 4.15 gives all significant configurations of the various factor interactions. While this table might appear fairly complex, a closer look at the configurations reveals that in essence there are only three results: the most important of these concerns all the types, i.e. the configurations which occur significantly more often than expected. As the only significant four-way factor configuration shows, all of these types involve British English main clause questions which have a stranded preposition heading a frequency PP. As pointed out above in section 4.3.1.4, these tokens all exhibit the idiosyncratically stranding discontinuous *how … for* idiom, which thus can be said to be much more deeply entrenched in British English. The other two significant configurations are antitypes which illustrate the rarity of (pied-piped) cleft relatives in Kenyan English and an avoidance of stranding in relative clauses (i.e. the ban on stranding with the obligatorily pied-piping PPs is even stronger there than in the other clause types; cf. section 4.4.2).

4.3.2 Logistic regression analysis

Before presenting the results of the Goldvarb analysis, it is important to survey the overall distribution of preposition placement in the various factor groups. When considering the raw frequencies of the individual groups, however, it should not be forgotten that these are merely indicative of potential effects. Due to the complex interaction of numerous factors only the statistical multivariate analysis can be regarded as identifying significant effects. Nevertheless, in order to highlight indicative trends in the following tables cells have been shaded (with light grey cells marking a preference for stranding, and dark grey indicating a preference for pied-piping).

Table 4.16 #*1 CLAUSE TYPES*

	ICE-GB			ICE-EA		
CLAUSE TYPE	stranded	pied-piped	total	stranded	pied-piped	total
Free	111 (98.2%)	2 (1.8%)	113	111 (98.2%)	2 (1.8%)	113
Question$_{\text{main clause}}$	88 (94.6%)	5 (5.4%)	93	39 (81.3%)	9 (18.8%)	48
Interrogative$_{\text{embedded}}$	52 (88.1%)	7 (11.9%)	59	44 (91.7%)	4 (8.3%)	48
RC	69 (13.6%)	439 (86.4%)	508	49 (15.9%)	260 (84.1%)	309
Cleft	8 (14.0%)	49 (86.0%)	57	3 (60.0%)	2 (40.0%)	5
Total	328 (39.5%)	502 (60.5%)	830	246 (47.0%)	277 (53.0%)	523

Starting with the first factor group, Table 4.16 gives an overview of preposition placement in the various clause types. As this shows, in both British and Kenyan English, free relatives, embedded interrogatives and main clause questions seem to favour stranding, while relative clauses exhibit a preference for pied-piping. The only notable difference between the two varieties concerns cleft relatives, which in the ICE-EA corpus slightly favour stranding (though it should be noted that this clause type only occurs fairly infrequently in the Kenyan corpus).

Next, Table 4.17 summarizes the distribution of the factors in group #2, i.e. the various different displaced elements. A first glance at the table might lead one to deduce that for both varieties there are some displaced elements which favour stranding (*what*, *who*, *how*, *where* and *whose*), while others clearly favour pied-piping (*which* and *whom*). However, it is especially the results from this group which will have to be interpreted with caution due to several interaction effects with factor group #2: NPs, for example, only occur in cleft relatives, which license no other displaced element (which thus accounts for the different effect of NP fillers in British and Kenyan English). In contrast to this, all other clause types exclusively require displaced *wh*-words. *What*, on the other hand, does not appear in relative clauses. Finally, the strong preference of pied-piping with *which* only seems to be a secondary effect of its distribution across clause types: in both corpora, the vast majority of *which*-example tokens occur in relative clauses (cf. the following (*which*-relative clause/overall *which*-tokens) ratios: ICE-GB: 403/404 = 99.8% pied-piped and 53/69 = 76.8% stranded; ICE-EA: 237/238 = 99.6% pied-piped and 36/37 = 97.3% stranded).

It will therefore only be possible to assess the precise effect of the factors in this group after the multivariate analysis. What is interesting about the results in Table 4.17, however, is that the ICE-GB corpus also contains one instance of a pied-piped *who*-example (4.83) and that both corpora include *whom*-tokens with a stranded preposition (4.84):

Table 4.17: #2 *DISPLACED ELEMENT*

	ICE-GB			ICE-EA		
DISPLACED ELEMENT	stranded	pied-piped	total	stranded	pied-piped	total
What	184 (96.3%)	7 (3.7%)	191	146 (94.8%)	8 (5.2%)	154
Who	22 (95.7%)	1 (4.3%)	23	2 (100%)	0 (0%)	2
How	14 (87.5%)	2 (12.5%)	16	7 (63.6%)	4 (36.4%)	11
Where	23 (82.1%)	5 (17.9%)	28	42 (85.7%)	7 (14.3%)	49
Whose	6 (66.7%)	3 (33.4%)	9	3 (60.0%)	2 (40.0%)	5
Which	69 (14.6%)	404 (85.4%)	473	37 (13.5%)	238 (86.5%)	275
NP	8 (14.0%)	49 (86.0%)	57	3 (60.0%)	2 (40.0%)	5
Whom	2 (6.1%)	31 (93.9%)	33	6 (27.3%)	16 (72.7%)	22
Total	328 (39.5%)	502 (60.5%)	830	246 (47.0%)	277 (53.0%)	523

(4.83) 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>

(4.84) a. Well you also I I I I think you need uh to see somebody frequently **who whom you can shout at** <ICE-GB:S1A-062 #110:1:A>

b. was announced by the Ministry of Culture in France and not by his wife Rodeka **whom he stayed with** for fifty-eight years <ICE-EA:S2B039K>

As mentioned in chapter 3, pied-piping with *who* is considered to be strongly disfavoured, sometimes even regarded as ungrammatical. Yet (4.83) seems perfectly acceptable, possibly due to the existence of overtly similar structures in *in situ wh*-questions such as *Who gave what to who?* (see section 3.1.2.1). On the other hand, (4.84a) shows that the speaker consciously changed a *who* to a *whom*-relativizer, indicating the grammaticality of preposition-stranding with *whom*. Finally, in section 3.6.3 it was pointed out that Kenyan secondary school text books sometimes appear to have an implicit preference for *whom* in stranded contexts, so that formal education might be partially responsible for the fact that structures like (4.84b) are even more frequent in ICE-EA than ICE-GB (despite the lower overall token size of the former corpus). Obviously, both phenomena – pied-piping with *who* and stranding with *whom* – are low-frequency phenomena, which makes it difficult to draw any definite conclusions about their grammaticality from corpus data alone. Yet, examples like (4.83) and (4.84) clearly warrant further experimental analysis (which will be presented in chapter 5).

The first variable whose results appear to differ more drastically across the two varieties is #3, TYPE OF XP CONTAINED IN (Table 4.18). In British English pied-piping is more frequent than stranding regardless of the XP

Table 4.18 *#3* TYPE OF *XP* CONTAINED IN

	ICE-GB			ICE-EA		
XP	stranded	pied-piped	total	stranded	pied-piped	total
VP	305 (40.3%)	451 (59.7%)	756	229 (48.1%)	247 (51.9%)	476
AdjP	13 (39.4%)	20 (60.6%)	33	14 (70.0%)	6 (30.0%)	20
NP	10 (24.4%)	31 (75.6%)	41	3 (11.1%)	24 (88.9%)	27
Total	328 (39.5%)	502 (60.5%)	830	246 (47.0%)	277 (53.0%)	523

which the PP was contained in. In the ICE-EA, however, AdjP-contained PPs seem to favour stranding, while NP-contained ones apparently lead to pied-piping. VP-contained PPs, which make up the majority of all ICE-EA tokens (476/523 ≅ 91.0%), appear equally often with either preposition placement variant. As always, however, the precise effect of these factors can only be discovered by the multivariate analysis.

Next, Table 4.19 presents the distribution of factors of group #4, TEXT TYPE, whose raw frequencies seem to exhibit the expected trend: the most informal text types (private dialogue in both corpora and private correspondence in ICE-GB) exhibit mainly stranded prepositions; the most formal ones (all printed and edited texts) indicate a strong preference for pied-piped prepositions. What is striking about Table 4.19 is that ICE-EA written edited texts contain many more stranded prepositions (31.1%) than the corresponding British English data (10.0%), which might indicate different copy-editing procedures. Apart from that, it was pointed out earlier (see section 2.2.1) that pragmatic reasons required the ICE-EA sampling scheme to slightly depart from the ICE-GB one, e.g. unscripted monologues are not part of the Kenyan corpus. On top of that, ICE-EA has two additional text types not found in ICE-GB, which exhibit contrary effects with respect to preposition placement: while stranding is the more common variant in the 'written-as-spoken' texts (i.e. spontaneous dialogue in relatively formal settings, namely the house of parliament and courtrooms), pied-piping is the more common choice in legal presentations (which are formal written manuscripts which are to be read out). Moreover, the distribution of the 'written-as-spoken' texts is roughly similar to that found in public dialogues, which in the ICE-EA corpus have a greater preference for stranded prepositions (76.5%) than in the ICE-GB corpus (55.2%; see Table 4.19). Note, however, that the slightly different types of sampled texts turned out to be unproblematic: as the statistical analysis showed, it was possible to significantly simplify this variable by collapsing the majority of factors with similar effects (see below).

Table 4.19 *#4 TEXT TYPE*

	ICE-GB			ICE-EA		
TEXT TYPE	stranded	pied-piped	total	stranded	pied-piped	total
Private dialogue	124 (95.4%)	6 (4.6%)	130	29 (80.6%)	7 (19.4%)	36
Private correspondence	14 (82.4%)	3 (17.6%)	17	6 (50.0%)	6 (50.0%)	12
Public dialogue	85 (55.2%)	69 (44.8%)	154	65 (76.5%)	20 (23.5%)	85
Written-as-spoken	NA	NA	NA	32 (71.1%)	13 (28.9%)	45
Legal presentations	NA	NA	NA	2 (14.3%)	12 (85.7%)	14
Unscripted monologue	54 (44.6%)	67 (55.4%)	121	NA	NA	NA
Business correspondence	7 (43.8%)	9 (56.3%)	16	2 (25.0%)	6 (75.0%)	8
Non-professional writing	7 (16.7%)	35 (83.3%)	42	9 (27.3%)	24 (72.7%)	33
Mixed	3 (15.0%)	17 (85.0%)	20	1 (11.1%)	8 (88.9%)	9
Scripted monologue	8 (10.8%)	66 (89.2%)	74	35 (48.6%)	37 (51.4%)	72
Printed / edited texts	26 (10.0%)	230 (90.0%)	256	65 (31.1%)	144 (68.9%)	209
Total	328 (39.5%)	502 (60.5%)	830	246 (47.0%)	277 (53.0%)	523

Finally, factor group #5 PP TYPE also displays the expected major effects. As can be seen in Table 4.20, lexically stored V-X-P idioms and prepositional verbs appear more often with stranded prepositions (with a frequency of 96.7%/92.3% and 63.2%/74.6% in ICE-GB/ICE-EA), while time and location/direction PP adjuncts co-occur more often with pied-piped prepositions (with a frequency of 98.6%/96.2% and 93.7%/97.8% in ICE-GB/ICE-EA). All other PPs align themselves somewhere on the cline between those two extremes. However, as pointed out in section 4.3.1.4, cause/reason/result, frequency, instrument and subcategorized PPs have an apparently even greater preference for pied-piping in the Kenyan data. Again, however, the statistical data analysis indicated that the factor group should be significantly simplified.

Next, in order to detect the statistically significant effects underlying the distribution of preposition placement in the two corpora the above five variables together with the factor group VARIETY (British English vs Kenyan English) were subjected to a single Goldvarb regression analysis.

During the first steps of this analysis, it turned out to be statistically justified to collapse the factor group TEXT TYPE into a three-way distinction of ‘informal’ (comprising private dialogues and private correspondence), ‘medium’ (the factors mixed, public dialogue, unscripted monologue and written-as-spoken) and ‘formal’ texts (business correspondence, legal presentations, non-professional writing, printed/edited texts and scripted

Table 4.20 #5 *PP* TYPE

	ICE-GB			ICE-EA		
XP	stranded	pied-piped	total	stranded	pied-piped	total
V-X-P	29 (96.7%)	1 (3.3%)	30	12 (92.3%)	1 (7.7%)	13
Prepositional verb	103 (63.2%)	60 (36.8%)	163	91 (74.6%)	31 (25.4%)	122
Movement	25 (58.1%)	18 (41.9%)	43	38 (63.3%)	22 (36.7%)	60
Optional PP complement	133 (48.5%)	141 (51.5%)	274	98 (66.7%)	49 (33.3%)	147
Accompaniment	5 (45.5%)	6 (54.5%)	11	1 (33.3%)	2 (66.7%)	3
Frequency	7 (20.6%)	27 (79.4%)	34	0 (0%)	14 (100%)	14
Cause/reason/result	6 (20.7%)	23 (79.3%)	29	0 (0%)	24 (100%)	24
Instrument	10 (19.2%)	42 (80.8%)	52	0 (0%)	49 (100%)	49
Obligatory complement	2 (18.2%)	9 (81.8%)	11	2 (16.7%)	10 (83.3%)	12
Subcategorized PP	1 (12.5%)	7 (87.5%)	8	0 (0%)	3 (100%)	3
Location / direction	5 (6.3%)	74 (93.7%)	79	1 (2.2%)	44 (97.8%)	45
Affected location	1 (4.5%)	21 (95.5%)	22	2 (40.0%)	3 (60.0%)	5
Time	1 (1.4%)	73 (98.6%)	74	1 (3.8%)	25 (96.2%)	26
Total	328 (39.5%)	502 (60.5%)	830	246 (47.0%)	277 (53.0%)	523

monologue).[9] This recoded factor group TEXT TYPE then gave rise to an interesting three-way interaction with the variables VARIETY and CLAUSE TYPE.[10] Figures 4.1 and 4.2 illustrate the nature of this interaction.

The graphs for both British and Kenyan English in Figure 4.1 show that preposition-stranding is strongly favoured in main clause and embedded questions as well as free relatives regardless of the level of formality. Even in those cases where there seems to be a minor formality effect (cf. the slight decrease of stranding across text types for embedded questions in British English and for main clause questions in Kenyan English), the most formal contexts still exhibit over 70 per cent stranding.

The situation is completely different with relative clauses and cleft sentences, which due to similar effects were pooled together. The upper graph of Figure 4.2 clearly demonstrates a formality effect in British English: preposition-stranding is the favoured choice in bound and cleft relative clauses in informal text types and pied-piping is dominant in medium and formal texts. In contrast to this, pied-piping is favoured in Kenyan English regardless of

[9] As for all recodings, a series of G^2-log likelihood tests were performed to ensure that simplifications did not lead to any statistically significant decrease in model fit (cf. Paolillo 2002: 140–2).

[10] A model with only these three variables had resulted in a rejected Fit: X-square(3) of 11.965 (p = 0.0089).

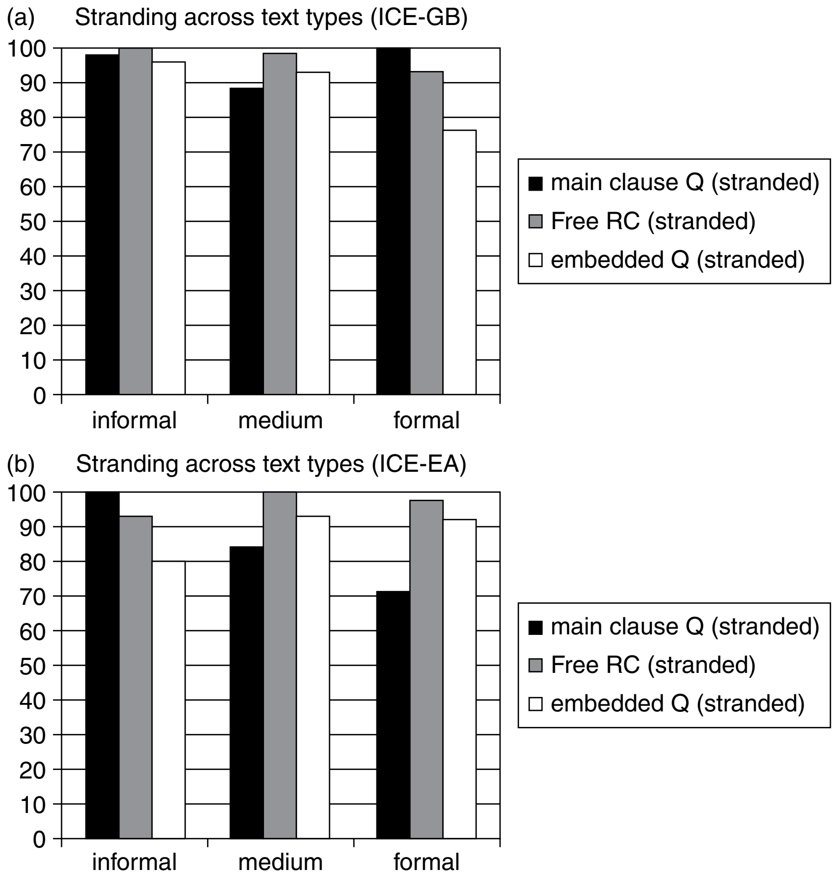

Figure 4.1 Preposition stranding in non-relative clauses across TEXT TYPE and VARIETY

the level of formality, though less formal text types at least tend to decrease this effect (cf. the lower graph of Figure 4.2).

In light of these findings, it was possible to capture this interaction by recoding the data in the following way: due to similar effects free relatives, main clause questions and embedded interrogatives were combined into a single factor ('Else'). The set of cleft-relatives and ordinary relative clauses was instead coded for both TEXT TYPE and VARIETY: the statistical analysis confirmed that the variable TEXT TYPE could be further simplified by collapsing it into two factors 'more formal' (the above medium and formal texts) and 'less formal' (the above informal group of private dialogue and private correspondence). In a next step these less formal clauses were then subdivided into Kenyan (less formal Kenyan clefts/relatives) and British English (less formal British clefts/relatives). Formal cleft relatives and relative clauses were subjected to the same procedure, also yielding individual

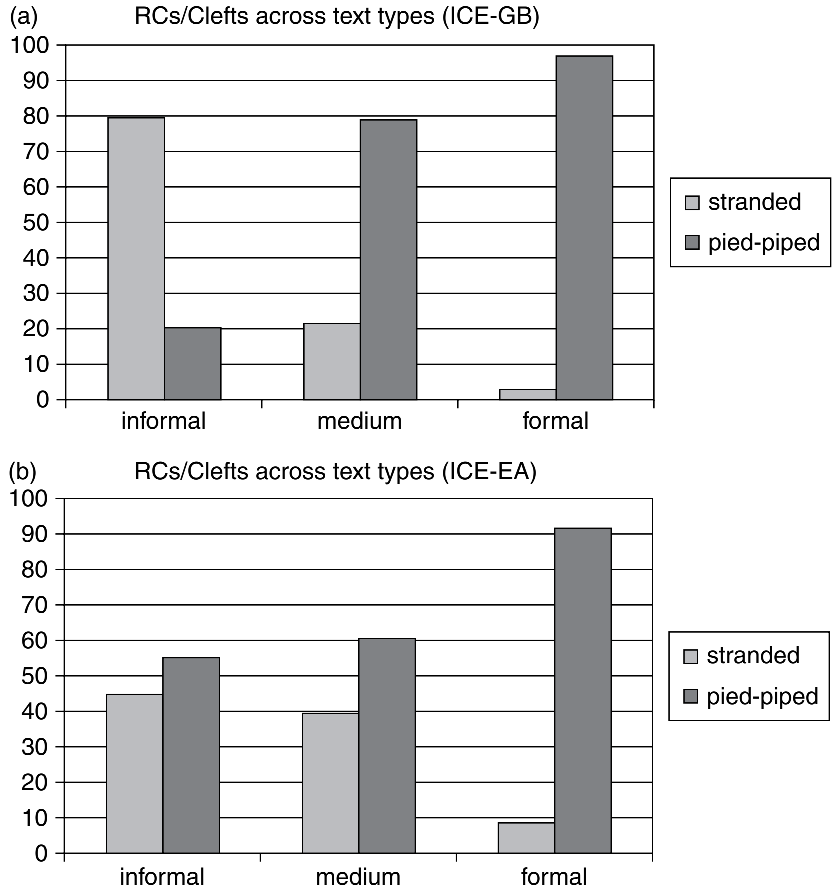

Figure 4.2 Preposition placement in relative/cleft clauses across TEXT TYPE and VARIETY

factors for Kenyan (more formal Kenyan clefts/relatives) and British English (more formal British clefts/relatives).

The final, most parsimonious model with the best fit for the data resulting from these recodings is given in Table 4.21. As the fit parameters indicate, the model in Table 4.21 can be considered a good fit for the data (for the full statistical evaluation see Appendix A.4.3.2).[11] Furthermore, it turns out that the only variable which did not turn out to have a significant effect was DISPLACED ELEMENT. As the logistic regression analysis showed, within the factor group DISPLACED ELEMENT only *whose* seemed to have an effect

[11] The binomial one-level analysis only identified one problematic cell whose expected frequency differed significantly from the actually observed one. As it turned out the only troublemaker was a cell with only one token, whose expected frequency was considerably lower than 5 (0.102), which violates one of the basic assumptions of chi-square tests and explains the apparently high Error value (while all other parameters signal a good model fit).

Table 4.21 *Final Goldvarb model for the variable ICE-EA and ICE-GB data*

Factor group (Significance)	Factor	Stranded/total (% stranding)	Goldvarb weight	Log odds
CLAUSE TYPE * TEXT TYPE *	Free RC Interrogative$_{embedded}$ Question$_{main\ clause}$	445 / 474 (93.9%)	0.957	2.731
VARIETY ($p < 0.001$)	Less formal**wh*-/cleft-RC (British English)	27 / 34 (79.4%)	0.916	2.025
	Less formal**wh*-/cleft-RC (Kenyan English)	9 / 20 (45.0%)	0.473	−0.475
	More formal**wh*-/cleft-RC (Kenyan English)	43 / 294 (14.6%)	0.192	−1.802
	More formal**wh*/cleft-RC (British English)	50 / 531 (9.4%)	0.108	−2.479
PP TYPE ($p < 0.001$)	Prepositional 'X' 'V-X-P' idioms	235 / 328 (71.6%)	0.830	1.446
	Optional complements Accompaniment Movement	300 / 538 (55.8%)	0.721	0.819
	Obligatory complements	4 / 23 (17.4%)	0.553	0.085
	Cause/reason/result Means/instrument Time Frequency Subcategorized PP Affected location Direction Location	35 / 464 (7.5%)	0.097	−2.360
TYPE OF XP CONTAINED IN ($p < 0.001$)	VP-contained AdjP-contained	561 / 1285 (43.7%)	0.514	0.569
	NP-contained	13 / 68 (19.1%)	0.254	−0.569

Fit: X-square(21) = 19.433, accepted, $p = 0.5250$[a] / Nagelkerke $R^2 = 0.784$
Cross-validation estimate of accuracy = 0.9

[a] As an anonymous reviewer pointed out, it is not always clear how Goldvarb exactly calculates these Fit: X-square values since a chi-square values of 19.433 at 21dfs should yield a p-value of e.g. $p = 0.5574$ (as calculated in R by pchisq(19.433, 21, lower.tail=F)). Yet, as the model's Nagelkerke R^2 as well as the cross-validation score show, this does not affect the validity of the model in Table 4.21.

on preposition placement: while *whose* favoured stranding across clause types and varieties (Goldvarb weight: 0.834; log odds: 0.814), all other fillers slightly favoured pied-piping (Goldvarb weight: 0.496; log odds: -0.814). Tentatively it could be speculated that stranding is strongly preferred with *whose* due to the complexity of the construction (cf. Hoffmann 2007b: 176). Take the following example:

(4.85) Pierce Inverarity, whose wealth she is to benefit from <ICE-GB:W1A-010 #16>

In (4.85), *whose* does not only function as a relativizer, it also occupies a determiner slot within the NP *whose wealth*. In order to interpret *whose* correctly the parser must therefore see inside the entire NP and a pied-piped preposition constitutes unnecessary intervening structure. Yet, this apparent effect of *whose* is based on a rather low token number of only fourteen instances. Moreover, the statistical analysis showed (cf. Appendix A.4.3.2) that the variable DISPLACED ELEMENT could be dropped and that the simpler model in Table 4.21 is actually a better fit for the data.

Turning now to the individual factor groups which Goldvarb gives as significantly influencing preposition placement, the interaction group VARIETY*CLAUSE TYPE*TEXT TYPE probably contains the most interesting results: as mentioned in section 4.2, free relative clauses do not categorically strand prepositions in all contexts. Nevertheless, these clauses have an almost categorical stranding effect (factor weight: 0.957 / log odds: 2.371). Even more importantly, embedded interrogatives and main clause questions pattern with free relatives in that they almost categorically favour stranding. Thus the functional constraint that these clauses should be introduced by a *wh*-word (see section 3.1) seems to have led to a strong stranding preference. In contrast to this, relative clauses generally favour pied-piping, which was claimed to be due to processing complexity: relative clauses are more complex to parse than, for example, questions and therefore favour pied-piping, which involves less processing effort than stranding (due to Hawkins's (2004) 'Avoid Competing Subcategorizors' and 'Valency Completeness' principles; cf. section 3.1.3). As the corpus studies showed, however, the situation is further complicated by the fact that in both varieties it is only relative clauses that are affected by formality effects: preposition pied-piping increases in relative clauses with increasing levels of formality. Yet, while this generalization holds for Kenyan and British English, only in the latter variety do informal contexts lead to stranding being favoured (cf. the factor weight of 0.916 / log odds of 2.025 for British English informal relative clauses). In Kenyan English the least formal contexts only result in a weaker pied-piping preference (cf. the factor weight of 0.473 / log odds of -0.475 for Kenyan English informal relative clauses). This indicates that the association of preposition-stranding and relative clauses in informal contexts is much stronger in British English (for the theoretical repercussions of all these findings see chapter 6).

The effects of the next factor group (PP TYPES) also include processing-based effects: as expected, prepositions which are lexically obligatorily associated with verbs (V-X-P idioms and prepositional verbs) strongly favour stranding (with a factor weight of 0.830 / log odds: 1.446), while adjunct-like PPs, including temporal and location adjunct PPs, favour pied-piping (inhibiting stranding with a factor weight of 0.097 / log odds -2.360). More interesting than this was the finding that accompaniment PPs and movement PPs pattern with optional PP complements (favouring stranding with a weight of 0.720 / log odds: 0.819). For movement PPs this 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. Their relatively strong preference for stranding, however, also implies a partly idiosyncratic lexical effect (see chapter 6). Another idiosyncratic effect is the strong pied-piping preference of subcategorized (e.g. *put something on the table / in the living room*) and affected location PPs (e.g. *sit on a chair*), since these can at least be said to be partly lexically associated with certain verbs. Nonetheless, these PP types pattern with the more adjunct-like PPs in the ICE data. Finally, obligatory complements are given as only mildly favouring stranding in Table 4.21 (factor weight: 0.553 / log odds: 0.085), which again is a result that cannot easily be accounted for by processing effects.

In contrast to this, the results of the factor group TYPE OF XP CONTAINED IN can fairly easily be attributed to processing constraints: NP-contained PPs disfavour stranding (factor weight: 0.254 / log odds: -0.569) since in these cases stranded prepositions are embedded in a constituent that itself is embedded in the main VP (cf. section 3.4). VP- and AdjP-contained PPs, on the other hand, do not exhibit a strong preference for either preposition placement variant (due to their weight of 0.514 / log odds: 0.569).

Now that the distribution of preposition placement across all clause types has been thoroughly analysed, the particular effects active in relative clauses will become the focus in the next chapter.

4.4 Relative clauses

As discussed in sections 3.1 and 3.1.2.2, relative clauses exhibit a number of interaction effects (e.g. the obligatory stranding constraint of *that* and Ø-relativizers, the categorical effect of non-finite relative clauses) which warrant a closer examination. The decision to further investigate the distribution of preposition placement in relative clauses is also supported by the fact that from a processing perspective, they constitute the most complex clausal context (see section 3.5). In addition to this, relative clauses are the most frequent clause type in which speakers seem to encounter preposition placement: in the ICE-GB study 66.7% (= 1180/1768) of all tokens were relative clauses (accounting for 488/985 = 49.5% of all stranded and 692/783 = 88.4% of all pied-piped tokens). The ICE-EA data display a similar distribution with 60.5% (=755/1247) of all instances occurring in relative clauses (making up 346/808 = 42.8% of all stranded and as many as 409/439 = 93.2% of all pied-piped tokens).

In the following, I will first turn to the analysis of the non-finite relative clause tokens in the two corpora (4.4.1) before discussing the status of the set of obligatorily pied-piping PP types (4.4.2). Then, I will investigate the tokens displaying variation (4.4.3), focusing especially on relative-clause-particular phenomena (such as the factor groups RESTRICTIVENESS and COMPLEXITY). Finally, all stranded finite relative clauses (including the categorical *that*- and Ø-tokens) will be subjected to an HCFA (4.4.4) in order to see whether these exhibit any significant factor combinations.

4.4.1 Non-finite relative clauses

In both ICE corpora non-finite relative clauses exhibit the expected categorical interaction effect: all Ø-relative clauses have a stranded preposition (4.86), while *wh*-relativizers only co-occur with pied-piped prepositions (4.87):

(4.86) a. And it's a it's **a very nice group to be working with** because I think everybody it's not too large <ICE-GB:S1A-002 #110:1:C>
b. But I always tell them please tell your boss tell your next mate tell your best friend that you're diabetic because those are the people who will save you and it's **nothing to be ashamed of** <ICE-EA:S1B041K:B>

(4.87) a. he celebrates within himself that he has at last found **someone with whom to dance** <ICE-EA:W1A004K>
b. So if you like I think that's probably you know quite **a good base on which to start** <ICE-GB:S1A-035 #70:1:B>

Non-finite relative clauses in which a PP is relativized on are a relatively rare phenomenon with only ninety-seven tokens in the ICE-GB corpus (seventy-one stranded Ø-tokens and twenty-six pied-piped *wh*-tokens) and sixty-one instances in the ICE-EA corpus (forty-seven stranded Ø-tokens and fourteen pied-piped *wh*-tokens). In order to investigate the categorical interaction effect of non-finite relative clauses more closely, the ICE data were subjected to an HCFA. Since the token size for this analysis was fairly low, it was decided to combine factors in a way that still allowed a thorough examination of the factor groups DISPLACED ELEMENT (*wh* vs Ø), TEXT TYPE and PP TYPE. For this, the factor groups were recoded as:

Table 4.22 *Non-finite RCs: HCFA configurations*

REL	TEXT	PP	VARIETY	Freq	Exp	Cont.chisq	O-E	P.adj.bin	Q
Ø	.	.	.	118	79.000	19.253	>	$p < 0.001$	0.494
wh				40	79.000	19.253	<	$p < 0.001$	0.494
.	Medium	.	.	67	52.667	3.901	>	$p < 0.05$	0.136
.	.	Complement	.	60	31.600	25.524	>	$p < 0.001$	0.225
.	.	$\text{PP}_{\text{obligatorily pied-piped}}$	.	8	31.600	17.625	<	$p < 0.001$	0.187
.	.	Be + P	.	11	31.600	13.429	<	$p < 0.001$	0.163
wh	.	$\text{PP}_{\text{obligatorily pied-piped}}$	.	8	2.025	17.626	>	$p < 0.05$	0.038
wh	.	Vprep	.	2	10.886	7.254	<	$p < 0.05$	0.060
Ø	.	$\text{PP}_{\text{obligatorily pied-piped}}$	.	0	5.975	5.975	<	$p < 0.05$	0.039
wh	Formal	Adjunct	.	10	2.711	19.597	>	$p < 0.05$	0.047
wh	Formal	Adjunct	KE	6	1.047	23.441	>	$p < 0.05$	0.032

a) TEXT TYPE: a three-way distinction of text types ('Informal', 'Medium', 'Formal') as in Figures 4.1 and 4.2;
b) PP TYPE: 'Vprep' (prepositional/V-X-P idiom verbs), 'Be + P', remaining 'Complement'-like PPs, remaining 'Adjunct'-like PPs (cf. Table 4.21) plus the factor '$\text{PP}_{\text{obligatorily pied-piped}}$' for the set of categorically pied-piping respect, manner, frequency and degree adjunct PPs.

This analysis gave the significant results shown in Table 4.22 (the full HCFA output can be found in Appendix A.4.4.1). The analysis of the individual variables presented in Table 4.22 shows that across both varieties, Ø-non-finite-relative clauses are significantly more frequent than those containing a *wh*-item. Furthermore, non-finite relative clauses occur significantly more often in texts of a medium level of formality (and not in informal or formal texts). Finally, prepositions which are lexically associated with verbs (the complement-like PPs) are more deeply entrenched in these constructions than all other PP types.

Turning to the remaining significant interaction configurations, the first thing to note is how the categorical effect of the different relativizers interacts with the choice of PP: since manner, respect, degree and frequency PPs ('X') obligatorily demand pied-piping, it is no surprise that these statistically favour *wh*-relativizers, which are compatible in terms of their preposition placement requirements. In contrast to this, these PPs do not occur with Ø-relativizers (frequency = 0) in the two corpora, a configuration which the HCFA identifies as a significant antitype. Furthermore, prepositional verbs and V-X-P idioms, which favour stranding, significantly disfavour *wh-* relativizers, which induce pied-piping. The final factor combination which the HCFA proves significant is pied-piping (due to a *wh*-relativizer) of adjunct PPs in formal texts, which again is hardly surprising given that the latter two factors also favoured pied-piping in the variable corpus data (see section 4.3). In addition to this, however, the statistical analysis also shows that it is especially in Kenyan English that this configuration is entrenched.

4.4.2 Categorically pied-piping PPs

In section 4.3 it was pointed out that respect, manner, degree and frequency PPs (with the exception of the lexicalized discontinuous structure *how … for* in interrogatives) obligatorily pied-pipe prepositions in those clause types that allow for variable preposition placement. As just seen in section 4.4.1, this effect even has immediate repercussions for the choice of relativizer in non-finite relative clauses. In addition to this, the statistical HCFA analyses of these PP types across British and Kenyan English revealed that they are particularly strongly associated with relative clauses (cf. Table 4.14). Next, I will investigate the effect of these PP types in relative clauses in more detail.

Corpora only yield positive data, which means that the absence of a phenomenon from a corpus does not entail its ungrammaticality. However, as I have argued above (section 4.3.1.4; also Hoffmann 2005, 2006), a closer inspection of corpora can sometimes help to separate data exhibiting apparent gaps from those which are the result of systematic effects. Take, for example, the distribution of PP types with a knockout effect in finite ICE-GB and ICE-EA relative clauses summarized in Table 4.23. All the PP types in Table 4.23 only occur with a pied-piped preposition in the finite *wh*-relative clause tokens (except for one stranded location adjunct token in ICE-EA). Yet, when other clausal contexts are taken into consideration it becomes clear that some of these PPs appear to exert a categorical pied-piping effect, while the absence of stranding with the other set of PPs is merely an accidental gap in the data.[12]

On the one hand there is a set of PP types which seem to demand obligatory pied-piping: in both corpora, respect (*people* ***for whom*** shopping is a drug), manner (*the ways* ***in which*** *they achieved it*), frequency/duration (*the frequency* ***with which*** *he saw her*) and degree adjuncts (*the extent* ***to which*** *it is true*) do not only pied-pipe categorically in finite *wh*-relative clauses, they also do not appear in either non-finite Ø- or finite *that*- and Ø-relative clauses, which would trigger stranding. I would argue that there is thus something like amounting negative data evidence in the corpus which supports the claim that stranding is impossible with these PP types.

[12] Note that this approach receives further support from experimental data that will be discussed in section 5.1: there it turned out that preposition-stranding with *wh*-, *that*- and Ø-relativizers was judged equally acceptable across all tested PP types by both British and Kenyan subjects.

Table 4.23 *PP types exhibiting knockout effects in ICE-GB RCs*

Variety	PP type	Finite *wh*-RC $P_{\text{Pied-piped}}$	Finite *wh*-RC P_{Stranded}	Finite non-*wh*-RC P_{Stranded}	Non-finite *wh*-RC $P_{\text{Pied-piped}}$	Non-finite non-*wh*-RC P_{Stranded}	Total
ICE-GB	Respect	119	0	0	1	0	120
	Manner	80	0	0	1	0	81
	Frequency	21	0	0	2	0	23
	Degree	28	0	0	1	0	29
	Subtotal	*248*	*0*	*0*	*5*	*0*	*253*
	Location	58	0	5	1	1	65
	Affected loc.	21	0	5	3	2	31
	Subcategorized	5	0	3	0	3	11
	Direction	4	0	2	0	0	6
	Subtotal	*88*	*0*	*15*	*4*	*6*	*113*
	Total (GB)	336	0	15	9	6	366
ICE-EA	Respect	77	0	0	0	0	77
	Manner	38	0	0	0	0	38
	Frequency	10	0	0	3	0	13
	Degree	17	0	0	0	0	17
	Subtotal	*142*	*0*	*0*	*3*	*0*	*145*
	Location	41	1	4	2	3	51
	Affected loc.	3	0	2	0	2	7
	Subcategorized	3	0	0	0	0	3
	Direction	3	0	1	0	0	4
	Subtotal	*50*	*1*	*7*	*2*	*5*	*65*
	Total (EA)	192	1	7	5	5	210

Subcategorized PPs (*the table* ***on which*** *she put it*), location (*the road* ***in which*** *the accident occurred*), affected location (*the bed* ***in which*** *they slept*), and direction PP adjuncts (*the room* ***into which*** *he walked*), on the other hand, appear in *that*- and (finite and non-finite) Ø-relative clauses. The knockout pied-piping effect of these PPs in *wh*-relative clauses thus looks like an accidental gap. (Note that in the light of the ICE-GB results, I take the absence of stranding with subcategorized PPs in the ICE-EA to be another accidental gap, caused by the low overall frequency of this phenomenon.)

Now one explanation of the above results might be blocking: with some of the above PPs it is possible to omit the preposition (e.g. *the ways they achieved it* or *the road the accident occurred*). The lack of stranded tokens with manner, degree, frequency and respect PPs in Table 4.23 might therefore be attributed to the fact that the preposition is regularly omitted in such contexts. Yet, such a view is far from unproblematic. First of all, Pesetsky (1998) analyses the omission of prepositions as the result of a constraint that allows the deletion of P + *wh*-sequences. In other words, the underlying structure of *the ways they achieved it* is *the ways in which they achieved it.*

Table 4.24 *Covarying-collexeme analysis of respect adjunct PPs*

VARIETY	word1	word2	frq.w1	frq.w2	obs.w1_2	exp.w1_2	relation	coll.strength	dec
BE	people	for	2	5	2	0.08	attraction	2.846	**
BE	framework	within	4	8	2	0.27	attraction	1.651	*
KE	circumstance	under	3	5	2	0.19	attraction	2.001	**

Since the constraint only applies to pied-piped prepositions, however, it does not explain the lack of stranded prepositions with the PPs in the above tables. Secondly, even if one assumes that stranded prepositions can also be omitted the problem remains that with many of the above PPs omission is not possible (cf. **people ~~for whom~~ shopping is a drug* or **the extent ~~to which~~ it is true*). Consequently, the effects of manner, respect, frequency and degree adjunct PPs cannot simply be accounted for by a potential blocking effect of preposition omission.

Instead, in Hoffmann (2005) I claimed that data such as those in the above tables indicate that there is a semantic constraint on preposition-stranding: in order for a stranded preposition to be interpretable the PP it heads must add thematic information to a predicate. Since respect, manner, frequency/duration and degree PP adjuncts do not add thematic information to a predicate (see section 4.3.1.1.), I argued that stranding with these PPs should be ungrammatical. Nevertheless, while the mounting negative data support such a conclusion, the data on which it rests are still subject to the negative data problem. Therefore, it was decided that further experimental data were needed to corroborate this hypothesis (see chapter 5).

Besides obligatory pied-piping with certain PPs, the potential idiosyncratic pied-piping effect of particular antecedents was mentioned in section 3.2.3.3. Since such effects are probably most deeply entrenched in constructions which generally pied-pipe prepositions, for each variety the set of finite *wh*-relative clauses with manner, respect, frequency and degree PPs was subjected to a covarying-collexeme analysis using Gries's (2004a) *Coll.analysis 3* script. In the following, I will present those antecedent + P structures which the Fisher-Yates exact test of the *Coll.analysis* analysis identified as significant. Note that I am only focusing on combinations which surface more than once in the corpus since single tokens with a significant effect might be the result of idiolectal preferences – a highly interesting claim but one which cannot be substantiated by corpus data alone. (For the full results of the analysis, see Appendix A.4.4.2.)

Starting with the set of respect adjunct PPs, the *Coll.analysis 3* script yielded the significant configurations for the British and the Kenyan English data shown in Table 4.24. As can be seen there, the antecedent + P structures *people for* (as in *people for whom the idea of going off-road involves driving into a puddle* <ICE-GB:S2A-055 #86:3:A>) and *framework within* (e.g. *the framework within which nationalism and politics could blossom* <ICE-GB:W2B-007 #38:1>) are identified as significantly associated in the ICE-GB data. In contrast to this, *circumstance under* comes out as the only entrenched Kenyan English antecedent + P collocation (cf. *the circumstances under which Ms. Imire died* <ICE-EA:S1BHN07K>).

Interestingly, for the set of manner adjunct PPs the covarying-collexeme analysis gives three antecedent + P sequences as significantly associated in the ICE-GB data, while there is no significant result for the ICE-EA data

Table 4.25 *Covarying-collexeme analysis of manner adjunct PPs*

VARIETY	word1	word2	frq.w1	frq.w2	obs.w1_2	exp.w1_2	relation	coll. strength	dec
BE	way	in	56	71	56	49.70	attraction	5.249	***
BE	ease	with	3	8	3	0.30	attraction	3.166	***
BE	speed	with	3	8	3	0.30	attraction	3.166	***
KE	way	in	22	35	22	20.26	attraction	1.178	n.s.

(cf. Table 4.25). In both corpora the antecedent *way* always co-occurs with *in* (as a comparison of the overall frequency of *way* (cf. column 'frq.w1' in Table 4.25) and the combination *way in* (cf. column 'obs.w1_2') shows). Moreover, these structures account for by far the majority of all manner tokens in both corpora (fifty-six out of eighty ICE-GB tokens and twenty-two out of thirty-eight ICE-EA tokens). Yet, while the particular combination of *way in* comes out as significant for British English, it does not for Kenyan English. The reason for this lies in the range of different prepositions employed in manner PPs: in Kenyan English *in* seems to be the standard preposition associated with manner adjunct PPs in relative clauses, regardless of the particular antecedent. In contrast to this, there is a greater range of prepositions used in British English in these constructions. This conclusion is also supported by the fact that not only the combination *way in* (e.g. *the ways in which the satire is achieved* <ICE-GB:S1B-014 #5:1:A>), but also *ease with* (*the ease with which the Saxons overran lowland England* <ICE-GB:W1A-001 #87:1>) and *speed with* (*the speed with which rainforests are being felled* <ICE-GB:W2B-028 #48:1>) turn out to be significant configurations in the ICE-GB data.

Next, Table 4.26 gives the results for the frequency PP tokens. For both varieties the combination *frequency with* comes out as a significant collocation (e.g. *the frequency with which the last digits appear* <ICE-GB:S1B-004 #230:1:A>; *frequency with which that quantity is consumed* <ICE-EA:W2A023K>). In addition to that, British English also has the entrenched combination *period for* (e.g. *the period for which your grant is payable* <ICE-GB:W2D-003 #53:1>).

Finally, relative clauses involving degree adjunct PPs also have entrenched antecedent + P sequences.For both British English and Kenyan English (see Table 4.27) the syntagm *extent to* is significantly associated (e.g. *the extent to which employers have increasingly accepted responsibilities in the field of short-term sickness* <ICE-GB:S1B-058 #72:1:C>; *the extent to which your proposals have been accepted by the Employer* <ICE-EA:W1B-BK13>). Moreover, *rate at* (e.g. *the rate at which the sediment in the uh on the floor of the Nile valley had risen with succeeding inundations* <ICE-GB:S2A-026 #38:1:A>), *degree to* (e.g. *the degree to which it provides a base for theories about psychological*

Table 4.26 *Covarying-collexeme analysis of frequency adjunct PPs*

VARIETY	word1	word2	frq.w1	frq.w2	obs.w1_2	exp.w1_2	relation	coll. strength	dec
BE	frequency	with	3	3	3	0.43	attraction	3.124	***
BE	period	for	4	2	2	0.38	attraction	1.544	*
KE	frequency	with	2	2	2	0.4	attraction	1.653	*

Table 4.27 *Covarying-collexeme analysis of degree adjunct PPs*

VARIETY	word1	word2	frq.w1	frq.w2	obs.w1_2	exp.w1_2	relation	coll. strength	dec
BE	rate	at	10	10	10	3.57	attraction	7.118	***
BE	extent	to	11	16	11	6.29	attraction	3.692	***
BE	amount	by	2	2	2	0.14	attraction	2.577	**
BE	degree	to	5	16	5	2.86	attraction	1.352	*
KE	extent	to	3	3	3	0.53	attraction	2.833	**

disorders <ICE-GB:W1A-007 #9:1>) and *amount by* (e.g. *the amount by which it absorbs light per unit of concentration* <ICE-GB:S2A-053 #49:1:A>) are significant antecedent–preposition combinations in the British English data.

As the preceding discussion has tried to show, there is a group of PPs, namely manner, respect, frequency and degree adjuncts, which seem to induce categorical pied-piping and are particularly associated with relative clauses in both British and Kenyan English. In addition to these effects, which will be further investigated in chapter 5, these tokens also exhibit partly variety-specific antecedent–preposition collocations, whose theoretical repercussions will be discussed in chapter 6.

4.4.3 Logistic regression analysis

After the distribution of categorical effects, the variable relative clause data were examined next. These consisted of 487 British English tokens (69 stranded prepositions and 418 pied-piped ones) and 301 Kenyan English tokens (50 stranded and 261 pied-piped ones). In addition to the factor groups used for the previous multivariate analyses, these relative clause data were coded for the factor groups RESTRICTIVENESS and COMPLEXITY (see section 4.1). Furthermore, due to the different effect of the level of formality in British and Kenyan English (see section 4.3), the interaction factor group TEXT TYPE*VARIETY was included in the analysis instead of the individual TEXT TYPE group (the factors of the interaction group being less formal

Kenyan relatives; more formal Kenyan relatives; less formal British relatives; more formal British relatives).

As an initial analysis of these relative clause tokens showed, this data set included the following knockout effects in the DISPLACED ELEMENT factor group: first of all, all nine *who*-relative clauses had a stranded preposition. Since the distribution of preposition placement with *who* and *whom* was further investigated experimentally (cf. chapter 5), it was therefore decided to combine them into a single factor for the Goldvarb analysis. Besides this, somewhat unexpectedly, *whose* only co-occurs with pied-piped prepositions in the less formal British and Kenyan data, while it favours stranding in more formal texts. In addition to this, *where* behaves differently in the British and Kenyan data: in the ICE-GB data *where* never yields a stranded preposition, in the ICE-EA data it exhibits the expected formality effect (i.e. less stranding in more formal texts). A closer inspection of the *whose*- and *where*-data reveals that these unexpected effects are due to the skewed distribution of these low frequency relativizers across the other factor groups: there are only four *where* tokens in British English, all of which appear pied-piped in more formal texts. In addition to that, the ICE-GB and the ICE-EA corpus only contain a single *whose*-token in less formal texts each, which in both cases happens to be pied-piped. These idiosyncratic effects might be an interesting topic for future research. For the present analysis, however, it was decided to exclude these tokens because of their low overall frequency and their resulting skewed distribution across other factors. This led to a reduced token set with 105 stranded (= 119 – 8 *whose* – 6 *where* tokens) and 655 pied-piped prepositions (= 669 – 4 *whose* – 10 *where* tokens).

The Goldvarb analysis of this reduced token set yielded models with an accepted model-fit, the most parsimonious of which is presented in Table 4.28 (see Appendix A.4.4.3 for details). The model in Table 4.28 meets the usual criteria of goodness of fit: it has an accepted Fit: X-square value and no cell with an expected frequency of 5 or higher exhibits an Error > 2. On top of that, the model has a moderate Nagelkerke R^2 value 0.454 and a cross-validation parameter of almost 90 per cent (cf. Appendix A.4.4.3.6).

Table 4.28 shows that the variables DISPLACED ELEMENT and COMPLEXITY did not turn out to have a significant effect on preposition placement in relative clauses. As Figure 4.3 indicates, however, the results concerning the latter group are nevertheless interesting.

In Figure 4.3, the factor group PP TYPES has been recoded into three factors (‘Vprep’, ‘Medium’, ‘Adjunct’), which correspond to the three groups of PPs identified by the logistic regression model in Table 4.28. On top of this, the factor group COMPLEXITY has been collapsed into a simple dichotomy of ‘simple’ (MCN < 3.5) and ‘complex’ (MCN ≥ 3.5; cf. sections 3.5.2 and 4.1). While an MCN value of 3.5 is admittedly a somewhat arbitrary cut-off point, Figure 4.3 reveals that it makes it possible to uncover an interesting interaction effect: if the preposition (‘Vprep’ = prepositional verbs

Table 4.28 *Final Goldvarb model for ICE-GB and ICE-EA RC data*

Factor group (Significance)	Factor	Stranded/total (% stranding)	Goldvarb weight	Log odds
PP TYPE ($p < 0.001$)	Prepositional 'X' V-X-P idioms	44 / 128 (34.4%)	0.895	1.636
	Obligatory complements Optional complements Accompaniment Movement	55 / 290 (19.0%)	0.758	0.637
	Cause/reason/result Means/instrument Time Frequency Subcategorized PP Affected location Direction Location	6 / 342 (1.8%)	0.146	−2.273
CLAUSE TYPE * TEXT TYPE * VARIETY ($p < 0.001$)	Less formal**wh*-RC (British English)	24 / 30 (80.0%)	0.987	2.860
	Less formal**wh*-RC (Kenyan English)	7 / 17 (41.2%)	0.860	0.360
	More formal**wh*-RC (Kenyan English)	34 / 267 (12.7%)	0.530	−1.334
	More formal**wh*-RC (British English)	40 / 446 (9.0%)	0.394	−1.886
RESTRICTIVENESS ($p < 0.001$)	Non-restrictive	48 / 183 (26.2%)	0.729	0.652
	Restrictive	57 / 577 (9.9%)	0.422	−0.652
TYPE OF XP CONTAINED IN ($p < 0.01$)	VP-contained AdjP-contained	100 / 704 (14.2%)	0.535	0.944
	NP-contained	5 / 51 (9.8%)	0.149	−0.944

Fit: X-square(26) = 25.473, accepted, $p = 0.2797$[a] / Nagelkerke $R^2 = 0.454$
Cross-validation estimate of accuracy = 0.891

[a] Again, as an anonymous reviewer pointed out, this Fit: X-square value is somewhat dubious: a chi-square value of 25.473 at 26dfs should yield a p-value of $p = 0.492$ (as calculated in R by pchisq(25.473, 26, lower.tail=F)), which implies an even better fit than estimated by Goldvarb.

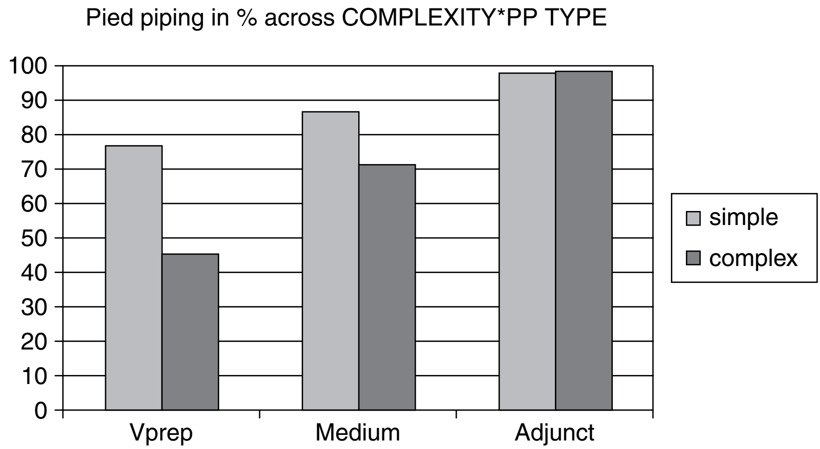

Figure 4.3 Pied-piping in relative clauses across COMPLEXITY and PP TYPE

and 'V-X-P' idioms) or the PP ('Medium' = obligatory and optional complements, accompaniment and movement PPs) is lexically associated with the main verb then increasing complexity leads to a decrease in pied-piping. The fact that increasing complexity thus favours stranding if there is a strong degree of lexical dependency between verb and preposition (cf. Hawkins 1999: 260, fn. 15) is also borne out by the logistic regression analysis (cf. A.4.3.3.3(i)): for these PP types, the preference for stranding in complex relative clauses ('Vprep' factor weight: 0.933 / log odds 2.634; 'Medium' factor weight: 0.805 / log odds: 1.418) is even higher than in simple clauses ('Vprep' factor weight: 0.771 / log odds: 1.214; 'Medium' factor weight: 0.639 / log odds: 0.571). In contrast to this, for the remaining, more adjunct-like PPs pied-piping is preferred regardless of the complexity of the relative clause (simple 'Adjunct' factor weight: 0.199 / log odds –1.393; complex 'Adjunct' factor weight: 0.197 / log odds: –1.405). These results are in line with earlier predictions (see Johansson and Geisler (1998: 76); Trotta (2000: 188)) and support the hypothesis that for adjunct PPs pied-piping is always easier to process, no matter what the complexity of the clause.

This interaction of COMPLEXITY and PP TYPE definitely warrants further investigation (see section 5.2). Yet, as the statistical analysis proved, once other complexity-related variables such as RESTRICTIVENESS and TYPE OF XP CONTAINED IN are taken into account, complexity can be dispensed with (see Appendix A.4.4.3.3). In the resulting most parsimonious model (Table 4.28), as expected from the overall corpus study in section 4.3, the formality of a text is shown to have a slightly different effect in British and Kenyan English: in British English relative clauses in less formal contexts strongly favour stranding (factor weight: 0.987 / log odds: 2.860), while more formal contexts favour pied-piping (inhibiting stranding with a weight of 0.399 /

log odds of –1.886). What is important is that Goldvarb factor weights must always be interpreted relative to those of the other factors within a group. It is for this reason that the factor weights for the level of formality differ from the ones obtained from the overall corpus study in section 4.3. This is most notable for the ICE-EA data: once all clause types are taken into account, it becomes clear that pied-piping with relative clauses is normally preferred regardless of the level of formality (cf. Table 4.21). If the focus is only on the relative clause tokens as in Table 4.28, however, the Goldvarb analysis simply indicates that while these generally favour pied-piping there is nevertheless a slight effect of the level of formality (cf. how the factor weight of 0.860 / log odds of 0.360 for less formal texts decrease to 0.530 / –1.334 in more formal Kenyan texts). Yet this effect is less pronounced in Kenyan English than in British English. For the assessment of the general status of the interaction of formality and relative clauses within the grammars of the two varieties, the results from the main corpus study take on greater importance, however, since these incorporate all variable clausal contexts.

With respect to the other groups identified as significant, the model includes the usual processing-based factors: PP types which are more closely associated with verbs strongly favour stranding (prepositional verbs and V-X-P idioms with a factor weight of 0.895 / log odds of 1.636 and the remaining complement-like PPs with a factor weight of 0.758 / log odds of 0.637), while more adjunct-like PPs favour pied-piping (inhibiting stranding with a factor weight of 0.146 / log odds of –2.273). In addition to this, NP-contained PPs also inhibit stranding with factor weight of 0.149 / log odds of –0.944 (therefore strongly favouring pied-piping).

The final group with a significant effect on preposition placement is a relative-clause-specific one: as can be seen in Table 4.28, non- restrictive *wh*-relatives favour preposition stranding (factor weight: 0.729 / log odds: 0.652), while restrictive ones slightly favour pied-piping (dispreferring stranding with a factor weight of 0.422 / log odds of –0.652). One reason why non-restrictive *wh*-relatives could be identified as a factor favouring stranding has to do with the ban on *that/Ø* in non-restrictive clauses: since non- restrictive *wh*-relativizers occur more frequently in contexts which in restrictive relative clauses favour both stranding and *that/Ø*, the factor non-restrictive itself might become interpreted as favouring stranding. In addition to this, non-restrictive clauses have weaker semantic ties with their antecedent (see, for instance, Olofsson 1981; Quirk *et al.* 1985). As is well known, these weaker semantic ties even have prosodic effects, since in speech non-restrictive relative clauses are often separated from their antecedent by a pause (Huddleston, Pullum and Peterson 2002). A pied-piped preposition might, however, be interpreted as establishing a closer relationship between the antecedent and the relative clause, fulfilling a kind of connective function. Note furthermore that the favouring stranding effect of non-restrictive relative clauses can also be seen as a matter of complexity: non-restrictive relative clauses are not necessary for the identification of the reference of the antecedent

Table 4.29 *HCFA: stranded finite RC data – individual variables*

RELATIVIZER	TEXT TYPE	PP TYPE	VARIETY	Freq	Exp	Cont.chisq	O-E	P.adj.bin	Q
Ø	.	.	.	332	102.143	517.259	>	$p < 0.001$	0.375
that	.	.	.	265	102.143	259.660	>	$p < 0.001$	0.266
.	Medium	.	.	354	238.333	56.135	>	$p < 0.001$	0.243
.	.	Complement	.	354	178.750	171.819	>	$p < 0.001$	0.327
.	.	Vprep	.	301	178.750	83.6087	>	$p < 0.001$	0.228

NP. Therefore the filler–gap identification process in non-restrictive relative clauses is less complex than in restrictive relative clauses (Hawkins 2004), which also accounts for the favouring stranding effect of the former.

4.4.4 HCFA: Stranded finite relative clauses

While the Goldvarb analysis allows the examination of the variable *wh*-tokens, the categorical *that*- and Ø-tokens had to be excluded due to their categorical knockout effect. Nevertheless, the question arises whether the latter relativizers license the same kind of stranded prepositions as their *wh*-counterparts. In other words, does preposition-stranding display the same kind of constraints regardless of the chosen relativizer (*wh*-, *that*- or Ø-) or are there any idiosyncratic effects?

In answer to this question, all finite stranded relative clause tokens were subjected to an HCFA. The data were analysed for the factor groups VARIETY (British vs Kenyan English), RELATIVIZER (*that*, Ø, *which*, *who*, *whom*, *whose*, *where*), TEXT TYPE ('Informal', 'Medium' and 'Formal') and PP TYPE ('Vprep' prepositional/V-X-P idiom verbs, 'Be + P', remaining 'Complement'-like PPs, remaining 'Adjunct'-like PPs; the factors of the latter two groups were thus identical to the ones employed in the analysis of the non-finite tokens in section 4.4.1). Tables 4.29 and 4.30 present the significant types identified by the HCFA (for the full analysis, see Appendix A.4.4.4).

Table 4.29 first of all gives the effect of the individual variables. As it turns out, *that*- and Ø-relativizers are identified as positively associated with stranded prepositions. Stranding is thus statistically more frequent with these relativizers than with the set of *wh*-words. In addition to this, stranded prepositions mostly occur in texts of a medium level of formality (though cf. below). Moreover, it is once again shown that lexically associated verb– preposition combinations favour stranding most across all types of relativizers.

Table 4.30 *HCFA: stranded finite RC data – interactions*

RELATIVIZER	TEXT TYPE	PP TYPE	VARIETY	Freq	Exp	Cont.chisq	O-E	P.adj.bin	Q
Ø	Formal			114	87.295	8.169	>	$p < 0.05$	0.043
	Formal	.	KE	122	78.3552	24.311	>	$p < 0.001$	0.069
	Informal	.	BE	134	100.897	10.861	>	$p < 0.01$	0.054
Ø	Formal	.	KE	72	36.383	34.867	>	$p < 0.001$	0.052
that	Medium	.	BE	105	76.520	10.600	>	$p < 0.05$	0.045
.	Informal	Complement	BE	84	49.954	23.203	>	$p < 0.001$	0.051
.	Formal	Vprep	KE	65	32.986	31.071	>	$p < 0.001$	0.047
Ø	Formal	Vprep	KE	38	15.317	33.594	>	$p < 0.001$	0.032
Ø	Informal	Complement	BE	43	23.196	16.909	>	$p < 0.05$	0.029

Table 4.30 then presents the interaction configurations which are positively associated. Within the set of finite stranded relative clauses, the HCFA exposes again another piece of evidence in support of the hypothesis that the level of formality plays a less important role in Kenyan English: while stranded prepositions in British English are associated with informal contexts, they are statistically more frequent in formal texts in the Kenyan data. Moreover, as the three- and four-way interactions show, it is particularly Ø-relativizers which surface significantly more often than expected in formal texts with prepositional verbs / V-X-P idioms in the ICE-EA data. In contrast to this, it is in informal texts that British English favours Ø-relativizers (in particular with complement-like PPs). Besides, in British English *that*-relative clauses with a stranded preposition are more frequent in texts of a medium level of formality.

4.5 Summary

In this chapter, the data from the ICE-EA and ICE-GB corpora were subjected to a number of statistical analyses. As it turned out, both Kenyan and British English are affected by several similar factors, yet both varieties also display some idiosyncratic effects.

Over and over again, British and Kenyan English were shown to be influenced by processing-related effects: the PP TYPES which are most closely associated with a particular verb, i.e. prepositional verbs and V-X-P idioms, were claimed to be most prone to stranding (cf. section 3.2). This claim was strongly supported by the corpus data: (1) these PPs are strongly associated with those clause types that induce categorical stranding (passives, non-*wh*-relatives, hollow and comparison clause; see section 4.2.3); (2) for all variable clause types, prepositional verbs and V-X-P idioms were always identified as favouring stranding most (see Table 4.21). In contrast to this, more adjunct-like PPs, such as temporal or locational adjunct PPs, exhibited the expected pied-piping preference: (1) in categorical stranding clauses they are less frequent than expected by chance (see section 4.2.3), and (2) in the variable data they always disfavour stranding the most (see e.g. section 4.3). Furthermore, the set of manner, respect, degree and frequency PPs was identified as leading to categorical pied-piping in British and Kenyan English (see section 4.3.1.4 and below for the idiosyncratic exceptional effect of *how … for*).

Besides the effect of various PP types, the influence of the factor groups TYPE OF XP CONTAINED IN and CLAUSE TYPE for all tokens with variable preposition placement as well as the influence of RESTRICTIVENESS on preposition-stranding in relative clauses were taken to be reflexes of processing complexity effects: NP-contained PP are more deeply embedded within a clause than VP- and AdjP-contained ones, which explains why stranding is disfavoured with the former type of phrase (as shown by the model in Table

4.21). In addition to that, non-restrictive relative clauses are easier to parse than restrictive ones, a reduction in processing cost that allows for preposition-stranding in these clauses. Free relative clauses, main clause questions and embedded interrogatives require even less processing effort than non-restrictive clauses (cf. sections 3.1.2, 3.5), which explains why stranding is favoured most strongly by these clause types (cf. Table 4.21).

The factor group CLAUSE TYPE, however, also exhibited a complex interaction effect with the group TEXT TYPE. As the multivariate analysis proved, only relative clauses are affected by the level of formality: free relative clauses, main clause questions and embedded interrogatives favoured stranding regardless of the formality of a text type. Relative clauses, on the other hand, had more pied-piped prepositions in more formal text types than in less formal ones. Moreover, this effect was slightly different in British and Kenyan English: while the ICE-GB data showed that less formal contexts led to stranding being favoured in relative clauses, in the ICE-EA data less formal texts merely caused a less pronounced pied-piping preference (see section 4.3.2). The conclusion that the level of formality exerts a weaker influence in Kenyan English received further support from the analysis of all finite relative clauses with a stranded preposition: there it turned out that these structures appear more frequently in more formal texts in Kenyan English and that it is Ø-relative clauses with prepositional verbs, especially, that are strongly associated with Kenyan English formal texts (see Table 4.30).

This different interaction effect of the level of formality and relative clauses was not the only idiosyncratic difference between British and Kenyan English. As the HCFA of the set of categorically stranding clauses indicated, non-*wh*-relatives are most strongly associated with British English (especially with complement-like PPs), while in Kenyan English it is passives that occur significantly more frequently (with prepositional verbs; see Table 4.10). Furthermore, free relative structures with an idiosyncratically stranding preposition such as *like* are more deeply entrenched in British English (Table 4.10). In fact, the particular collocation *be like* also turned out to be disfavoured in Kenyan English (cf. Table 4.12).

Next, idiosyncratic, variety-specific effects also surfaced in the set of categorically pied-piping manner, respect, frequency and degree PPs: only the ICE-GB data, for example, contained the lexically-stored *how… for* structure which licenses stranded prepositions with frequency PPs, which otherwise exhibit a categorically stranding constraint (a result which an HCFA proves significant; see Table 4.15). In addition to this, the covarying- collexeme analyses of these tokens yielded a number of different entrenched antecedent + P sequences for British and Kenyan English (see section 4.4.2).

All of these results will obviously have important repercussions for the discussion of the status of preposition-stranding and pied-piping in the mental grammars of speakers of British and Kenyan English. There are, however, still a few questions that required further experimental analysis: the corpus data, for example, had shown that *that*- and *Ø*-relative clauses obligatorily cause stranding, while manner, respect, frequency and degree PPs lead to obligatory pied-piping. Yet what is the ontological status of these two constraints? Does a violation of these constraints result in structures that are equally unacceptable? Furthermore, the precise effect of *who* and *whom* on preposition placement could not be established on the basis of the corpus data alone. Again, further experimental evidence was needed in order to investigate these factors in more detail. Finally, the factor group COMPLEXITY also required a further investigation. It was therefore decided to address all of these issues in a set of three introspection experiments, whose results will be the focus of the next chapter.

5 Evidence II: Experimental results

As the last chapter has shown, the examination of the Kenyan and British English components of the ICE corpus project already revealed a great number of processing-based as well as variety-specific constraints on preposition placement. Due to the inherent negative data problem of corpus evidence it was necessary, however, to corroborate some of the findings with experimental evidence. It was therefore decided to investigate the following issues in a series of introspection experiments using the Magnitude Estimation method:

(i) the status of pied-piping with *that*- and Ø-relativizers,
(ii) the status of pied-piping with the *wh*-words *who* and *whom*,
(iii) the alleged ungrammaticality of preposition stranding with certain PP types such as manner and frequency adjuncts,
(iv) the influence of syntactic complexity, and
(v) the acceptability of doubled preposition structures (*the place in which I live in*), a non-standard phenomenon for which it had only been possible to retrieve tokens from the ICE-EA corpus (cf. section 4.1).

The first experiment investigated issues (i) and (iii) in simple relative clauses (section 5.1), while the second one examined issues (i), (ii), (iv) and (v) in relative clauses of varying syntactic complexity (5.2). Finally, experiment number three focused on (ii), (iii) and (v) but this time tested these factors in interrogative clauses (5.3).

5.1 Preposition placement in simple relative clauses

The first experiment design crossed the following factors: PREPOSITION PLACEMENT (two levels: stranded; pied-piped), RELATIVIZER (three levels: *wh*-; *that;* Ø) and PP TYPE (three levels: prepositional verbs; temporal/locative sentence adjuncts; manner-degree/frequency-duration adjuncts). The factor RELATIVIZER allowed the testing of the hypothesis that preposition-stranding is equally acceptable with all three types of relativizers. Moreover, it was investigated whether pied-piping is equally ungrammatical in *that*- or Ø-relative clauses, or whether P + *that*-structures are in fact more similar to P + *wh*-sequences.

In addition to this, the experiment tested the claim that the distribution of the PP types in Table 4.23 permits one to distinguish categorical from accidental pied-piping effects. Therefore allegedly categorical PP types, namely manner/degree (e.g. *the way in which he achieved his goal*) and frequency/duration (e.g. *the frequency with which earthquakes occur*) adjuncts, were contrasted with allegedly accidentally pied-piping ones, namely locative and temporal sentence adjuncts (e.g. *an island on which he found gold*; *the day on which James arrived*). Finally, as a point of reference, this factor also included prepositional verbs (e.g. *I know the man on whom Jane relied.*), with which stranding is perfectly grammatical.

This design resulted in PREPOSITION PLACEMENT × RELATIVIZER × PP TYPE = 2 × 3 × 3 = 18 cells. Following Cowart (1997), every subject was exposed to all conditions, but never with the same lexical material (cf. section 2.3.1). Thus, for each of the three PP TYPE conditions, six different lexicalizations for every PREPOSITION PLACEMENT × RELATIVIZER factor combination were used. The different lexicalizations for each level of the factor PP TYPE are given in (5.1)–(5.3):

(5.1) **Prepositional verbs**:
- a. I know the man (**on**)$_1$ __ Jane relied (**on**)$_2$.
- b. Jennifer never calls the groupies (**with**)$_1$ __ she sleeps (**with**)$_2$.
- c. You wouldn't believe the things (**at**)$_1$ __ Bill laughs (**at**)$_2$.
- d. Sally fancies the guy (**about**)$_1$ __ Steve talked (**about**)$_2$.
- e. Brad did something (**for**)$_1$ __ he apologised (**for**)$_2$.
- f. Sarah never achieved the fame (**of**)$_1$ __ she dreamt (**of**)$_2$.

(5.2) **Locational/temporal adjunct PPs**:
- a. Poirot inspected the room (**in**)$_1$ __ the murder had taken place (**in**)$_2$.
- b. They stopped at a bar (**at**)$_1$ __ they enjoyed a few cocktails (**at**)$_2$.
- c. Matt retired to an island (**on**)$_1$ __ he found gold (**on**)$_2$.
- d. I forgot the day (**on**)$_1$ __ James arrived (**on**)$_2$.
- e. He was born in the year (**in**)$_1$ __ Elvis died (**in**)$_2$.
- f. She asked for the time (**at**)$_1$ __ the party started (**at**)$_2$.

(5.3) **Manner-degree/frequency-duration adjunct PPs**:
- a. Jack was surprised by the precision (**with**)$_1$ __ Ben worked (**with**)$_2$.
- b. I am not concerned with the way (**in**)$_1$ __ he achieved his goal (**in**)$_2$.
- c. His competitors couldn't believe the ease (**with**)$_1$ __ he'd won (**with**)$_2$.
- d. They attended a service (**during**)$_1$ __ they were not allowed to sit (**during**)$_2$.
- e. Bill told us about the frequency earthquakes (**with**)$_1$ __ occurred (**with**)$_2$.
- f. There have been several occasions (**on**)$_1$ __ Kelly fainted (**on**)$_2$.

For all the sentences in (5.1)–(5.3) all six possible PREPOSITION PLACEMENT × RELATIVIZER factor combinations were created (i.e. P_1 + *wh*- / P_1 + *that* / P_1 + Ø / *wh*- + P_2 / *that* + P_2 / Ø + P_2). The resulting total of 108 stimuli was then divided into six material sets of 18 stimuli by placing the items in Latin squares (Keller 2000; Keller and Alexopoulou 2005). In other words, the stimuli were counterbalanced as outlined in section 2.3.1 so that every subject encountered all PREPOSITION PLACEMENT × RELATIVIZER factor combinations for each of the three PP TYPES, but never saw any of the lexicalizations in (5.1)–(5.3) more than once. (Material set 1, for example, contained *I know the man on whom Jane relied* and *Sally fancies the guy who Steve talked about*, while material set 2 included *I know the man who Jane relied on* and *Sally fancies the guy about whom Steve talked*).

Furthermore, thirty-six relative clauses from the ICE-GB corpus were included as fillers, yielding a filler:stimuli ratio of 2:1. Eighteen of these fillers were manipulated to exhibit the following ungrammatical phenomena: six fillers with word order violations (5.4.a), six subject contact clauses (5.4b), six with subject–verb agreement errors (5.4c):

(5.4) **Ungrammatical filler sentences**:
- a. That's a tape I sent them that **done** ↔ **I've** myself
  (word order violation; original source: <ICE-GB:S1A-033 074>)
- b. There was lots of activity ~~**that**~~ goes on there
  (subject contact clause; original source: <ICE-GB:S1A-004 #067>)
- c. There are so **many people** who need**s** physiotherapy
  (subject–verb agreement error; original source: <ICE-GB:S1A-003 #027>)

The full set of fillers (together with material set 1) can be found in Appendix A.5.1.4.

As discussed in section 2.3.2, the method used in all experiments was Magnitude Estimation (Bard *et al.* 1996; Keller 2000). Thus subjects were asked to give numerical judgements on sentences proportional to a constant reference sentence. The experiment itself was conducted using the WebExp software (Keller *et al.* 1998). The experimental data were then normalized by transformation to z-scores and subjected to repeated measures analyses of variance (ANOVAs) by subjects and by items. The filler scores were obviously not part of the repeated measures ANOVA, but occasionally it became necessary to compare these statistically to the effect of selected experimental stimuli via a set of t-tests (using Bonferroni-corrected p-values; the significance threshold of 0.05 was thus always adjusted by the overall number of t-tests per experiment). Next I shall present the results for the British speakers before turning to those of the Kenyan speakers.

5.1.1 British English

Thirty-six native speakers of British English (eighteen female, eighteen male; age 17–64) were recruited for an online sentence acceptability experiment by personal invitation (students and lecturers of the universities of Edinburgh and Central Lancashire) as well as by a posting in an online chat room (www.thestudentroom.co.uk). The experiment ran from 13 January to 24 August, until a sufficient number of subjects had participated (i.e. an equal number of participants had filled in all six material set versions). A repeated measures ANOVA of these data (see Appendix A.5.1.1) then yielded significant main effects of

- PREPOSITION PLACEMENT:
  $F_1(1{,}33) = 4.536$, $p < 0.05$, $\eta^2 = 0.08$
  $F_2(1{,}5) = 32.261$, $p < 0.01$, $\eta^2 = 0.11$
- RELATIVIZER:
  $F_1(2{,}66) = 17.149$, $p < 0.001$, $\eta^2 = 0.21$
  $F_2(2{,}10) = 38.783$, $p < 0.001$, $\eta^2 = 0.29$
- PP TYPE (only significant by items):
  $F_1(2{,}66) = 0.997$, $p > 0.30$, $\eta^2 = 0.02$
  $F_2(2{,}10) = 30.281$, $p < 0.001$, $\eta^2 = 0.03$

as well as the interactions of

- PREPOSITION PLACEMENT*RELATIVIZER:
  $F_1(2{,}66) = 9.740$, $p < 0.001$, $\eta^2 = 0.20$
  $F_2(2{,}10) = 78.271$, $p < 0.001$, $\eta^2 = 0.27$
- PREPOSITION PLACEMENT*PP TYPE.
  $F_1(2{,}66) = 4.217$, $p < 0.02$, $\eta^2 = 0.08$
  $F_2(2{,}10) = 20.075$, $p < 0.001$, $\eta^2 = 0.11$

In contrast to this, the between-subject factors AGE ($F(1{,}33) = 2.760$, $p > 0.10$) and GENDER ($F(1{,}33) = 1.495$, $p > 0.20$) failed to reach significance.

In order to investigate the two interaction effects, a post-hoc Tukey test was conducted which gave the results in Table 5.1 and Table 5.2. (Note that numerical Magnitude Estimation judgements allow the calculation of means for particular factor combinations. Using these, the tables give the difference between means, with negative results therefore indicating that factor (II) has a higher mean than factor (I).)

As Table 5.1 shows, while all three types of relativizers are equally acceptable with preposition-stranding, the *wh*-relativizers are significantly more acceptable with pied-piping than either *that* or Ø (with *that* receiving slightly better scores than Ø). Table 5.2 then indicates that pied-piping with manner/frequency adjunct PPs receives significantly better scores than with prepositional verbs. Moreover, the stranded data exhibit a clear cline of acceptability: prepositional verbs are judged significantly better than locational/temporal adjunct PPs, which in turn are better than manner/frequency adjuncts.

In order to facilitate the interpretation of these results, in Figures 5.1 and 5.2 the mean judgements with standard error bars are plotted for pied-piping and stranding across RELATIVIZERS and PP TYPES. As can be seen, the two main effects of PREPOSITION PLACEMENT and RELATIVIZER can actually be attributed to the PREPOSITION PLACEMENT*RELATIVIZER interaction

Table 5.1 *Result of post-hoc Tukey test for* PREPOSITION PLACEMENT*RELATIVIZER *interaction*

Dependent variable	(I) Relativizer	(II) Relativizer	Mean difference (I–II)	Significance
$P_{pied\text{-}piped}$	*wh-*	*that*	1.411	$p < 0.001$
		Ø	1.704	$p < 0.001$
	that	*wh-*	−1.411	$p < 0.001$
		Ø	0.293	$p < 0.010$
	Ø	*wh-*	−1.704	$p < 0.001$
		that	−0.293	$p < 0.010$
$P_{stranded}$	*wh-*	*that*	0.057	ns
		Ø	0.014	ns
	that	*wh-*	−0.057	ns
		Ø	−0.044	ns
	Ø	*wh-*	−0.014	ns
		that	0.044	ns

Rows with significant effects are shaded in grey.

Table 5.2 *Result of post-hoc Tukey test for* PREPOSITION PLACEMENT*PP TYPE *interaction*

Dependent variable	(I) PP type	(II) PP type	Mean difference	Significance
$P_{pied\text{-}piped}$	$V_{prepositional}$	$Adjunct_{loc/temp}$	−0.086	ns
		$Adjunct_{man/frequ}$	−0.277	$p < 0.010$
	$Adjunct_{loc/temp}$	$V_{prepositional}$	0.086	ns
		$Adjunct_{man/frequ}$	−0.191	ns
	$Adjunct_{man/frequ}$	$V_{prepositional}$	0.277	$p < 0.010$
		$Adjunct_{loc/temp}$	0.191	ns
$P_{stranded}$	$V_{prepositional}$	$Adjunct_{loc/temp}$	0.418	$p < 0.001$
		$Adjunct_{man/frequ}$	0.868	$p < 0.001$
	$Adjunct_{loc/temp}$	$V_{prepositional}$	−0.418	$p < 0.001$
		$Adjunct_{man/frequ}$	0.450	$p < 0.001$
	$Adjunct_{man/degree}$	$V_{prepositional}$	−0.868	$p < 0.001$
		$Adjunct_{loc/temp}$	−0.450	$p < 0.001$

Rows with significant effects are shaded in grey.

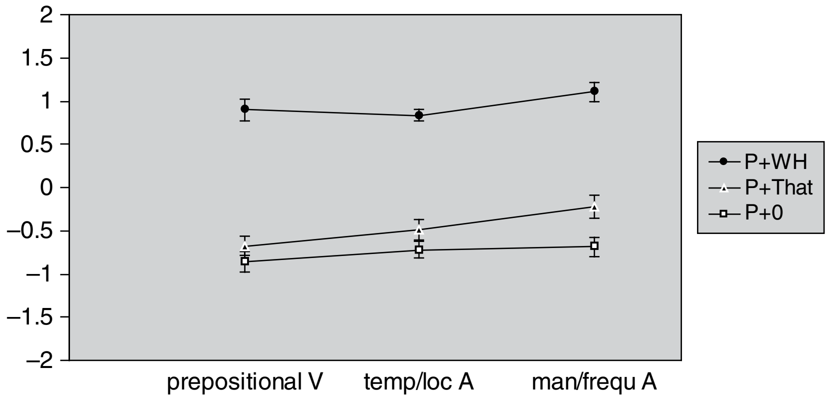

Figure 5.1 *Pied-piping across* RELATIVIZERS *and PP* TYPES *(British English speakers)*

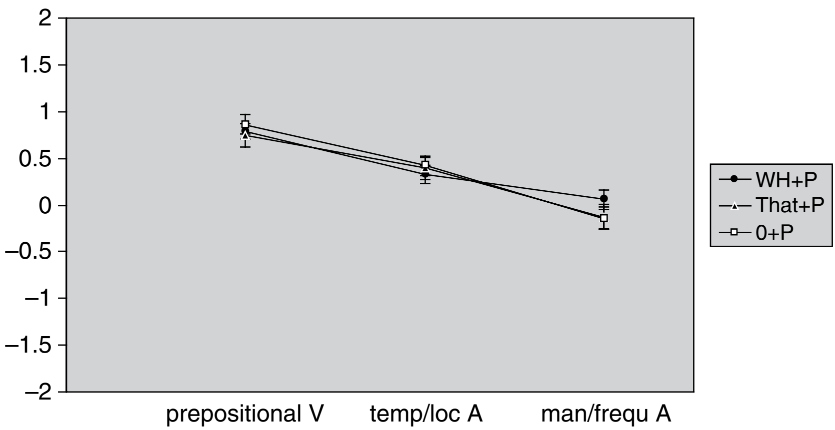

Figure 5.2 *Stranding across* RELATIVIZERS *and PP* TYPES *(British English speakers)*

effect, i.e. the ban on pied-piping with *that-* and Ø-relativizers: while the judgements of pied-piping with *wh*-relativizers across all PP types are significantly higher than with *that-* and Ø-relativizers (see Figure 5.1), all three relativizers are equally acceptable in relative clauses with preposition stranding (cf. Figure 5.2). The ungrammaticality of pied-piping with *that-* and Ø-relativizers was expected in the light of the absence of these constructions in the ICE-GB corpus and native-speaker introspection. However, the experiment also supported the hypothesis that across different PP types stranding is equally acceptable for the different relativizers. This consequently supports the use of *that-* and Ø-tokens as

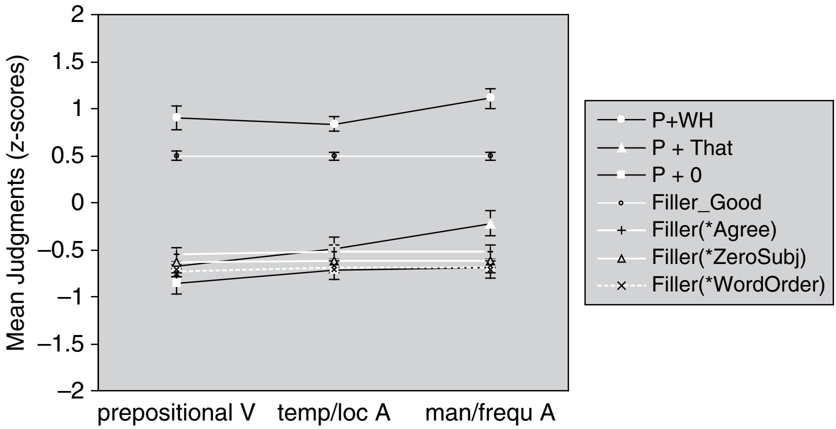

Figure 5.3 *Pied-piping across* RELATIVIZERS *and PP* TYPES *compared with fillers (British English speakers)*

intra-corpus corroborating evidence for the evaluation of the categorical effects of the *wh*-manner, degree, respect and frequency PP tokens in Table 4.23. Furthermore, this effect also helps to illustrate another advantage of carefully designed introspection experiments: the absence of a phenomenon in a corpus might indicate its ungrammaticality; subtle differences in judgements of such phenomena, however, cannot only corroborate such findings but, additionally, might reveal degrees of ungrammaticality (Kempen and Harbusch 2005; Sorace and Keller 2005). Now Magnitude Estimation yields gradient judgements, but this of course does not automatically entail that these gradient differences in acceptability entail gradient differences in grammaticality. However, since all judgements in Magnitude Estimation experiments are always relative, contrasting the judgements of various constructions gives extremely insightful results. Take for example Figure 5.3, which plots the mean judgements together with standard error bars of pied-piping across RELATIVIZERS and PP TYPES (just like Figure 5.1), but also gives the mean judgements for the various types of filler sentences used in the experiment. As Figure 5.3 shows, pied-piping with *that* and Ø-relativizers is judged as being considerably worse than pied-piping with *wh*-relativizers or the grammatical fillers. Instead, the two constructions pattern at the very end of the acceptability cline along with the ungrammatical fillers. Following Sorace and Keller (2005), this can be taken as an indication of the fact that pied-piping with *that*- and Ø-relativizers violates a hard grammatical constraint.

An additional point to note about the pied-piped judgements (cf. Table 5.2) is the fact that the manner/frequency adjunct tokens received the highest judgements out of all stimuli (experimentals and fillers). This might be taken as another indication of the lexicalization of such strings as *way in*

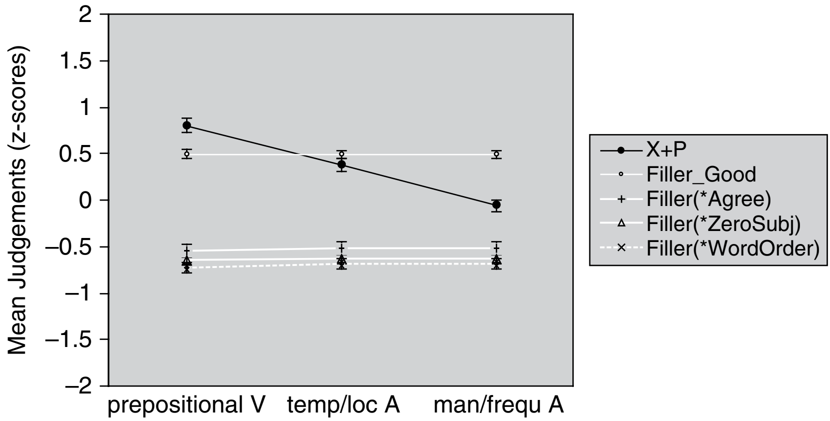

Figure 5.4 *Stranding means for all* RELATIVIZERS *across PP* TYPES *compared with fillers (British English speakers)*

which or *degree to which* as complex manner/frequency adjunct relativizers (on a par with free relativizers such as *what* 'that which' in e.g. *He gave me what I wanted*).

While pied-piping with *that* and Ø were both treated as violations of hard grammatical constraints by subjects, Figure 5.2 already indicates that the PREPOSITION PLACEMENT*PP TYPE interaction includes the violation of a soft constraint. As Figure 5.4 shows, preposition-stranding with prepositional verbs is judged better than with the other two PP-type contexts. The stranded temporal/location adjunct PPs in turn are judged better than the manner/frequency adjunct tokens.

Note that the cline of acceptability in Figure 5.4 corroborates the hypothesis based on the corpus data: a preposition can only be stranded if it heads a PP which contributes interpretable thematic information to the predicate. Therefore it comes as no surprise that stranded manner/frequency adjunct tokens are the only PP context which is judged significantly worse than the grammatical fillers ($t(35) = -5.905$, $p < 0.001$).[1] Yet compared with the remaining filler stimuli, it is also important to see that stranding with manner/frequency adjuncts receives better judgements than the set of ungrammatical fillers. Preposition-stranding can thus be considered a soft grammatical constraint (Sorace and Keller 2005).

[1] Since, all in all, ten such t-tests were calculated, the Bonferroni-corrected p-value for these was $0.05/10 = 0.005$. As it turned out, temporal/location adjuncts are judged as good as the grammatical fillers ($t(35) = -1.349$, $p > 0.18$), while prepositional verbs are considered better than the grammatical fillers ($t(35) = 3.728$, $p < 0.005$). The latter effect can be explained by the fact that prepositional verbs such as *rely on* or *talk to* are stored as complex lexical items, which facilitates the interpretation of such V-P structures.

As the above shows, the independent effects of PREPOSITION PLACEMENT, RELATIVIZER and PP TYPE (by items only) can be accounted for by the PREPOSITION PLACEMENT*RELATIVIZER and PREPOSITION PLACEMENT*X-PP-RELATIONSHIP interactions. Retaining only the two interactions for a further ANOVA analysis then showed that these two factors account for 61 percent of the total by-subjects variability of the experiment (PREPOSITION PLACEMENT*RELATIVIZER: $\eta^2 = 0.43$ + PREPOSITION PLACEMENT*X-PP-RELATIONSHIP: $\eta^2 = 0.18$) and 89 percent of the by-items total variability (PREPOSITION PLACEMENT*RELATIVIZER $\eta^2 = 0.63$ + PREPOSITION PLACEMENT*X-PP-RELATIONSHIP $\eta^2 = 0.26$).

5.1.2 Kenyan English

Due to technical problems none of the experiments could be completed online by Kenyan speakers. During a field trip to Kenya in September 2006, thirty-six speakers of Kenyan English (twenty-seven female, nine male; age 21–28) were therefore recruited at the University of Nairobi to fill in a printed questionnaire version of the experiment (which also had been gener ated by the WebExp software). While it would have been interesting to get equal numbers of speakers of the two main language families (Bantu and Nilotic), the large majority of subjects turned out to be speakers of a Bantu language (mainly Kikuyus, who make up the majority of students at the University of Nairobi). Only five out of thirty-six subjects were native speakers of a Nilotic language (four Dholuo-speakers and one Kalenjin-speaker). As a result the potential effect of first languages had to be left for future studies to investigate.

A repeated measures ANOVA of the Kenyan English data (see Appendix A.5.1.2) then did not give any effect for the between-subject variables AGE ($F(5,23) = 1.610$, $p > 0.20$) and GENDER ($F(2,23) = 1.676$, $p > 0.20$). The set of within-subject variables, however, yielded significant main effects of

- PREPOSITION PLACEMENT:
  $F_1(1,35) = 40.916$, $p < 0.001$, $\eta^2 = 0.06$
  $F_2(1,5) = 27.738$, $p < 0.005$, $\eta^2 = 0.17$
- RELATIVIZER:
  $F_1(2,70) = 17.413$, $p < 0.001$, $\eta^2 = 0.05$
  $F_2(2,10) = 10.939$, $p < 0.005$, $\eta^2 = 0.12$
- PP-TYPE:
  $F_1(2,70) = 4.086$, $p < 0.05$, $\eta^2 = 0.01$
  $F_2(2,10) = 4.900$, $p < 0.05$, $\eta^2 = 0.03$

as well as the interactions of

- PREPOSITION PLACEMENT*RELATIVIZER:
  $F_1(2,70) = 28.163$, $p < 0.001$, $\eta^2 = 0.09$
  $F_2(2,10) = 28.350$, $p < 0.001$, $\eta^2 = 0.18$

Table 5.3 *Result of post-hoc Tukey test for* PREPOSITION PLACEMENT*RELATIVIZER *interaction*

Dependent variable	(I) Relativizer	(II) Relativizer	Mean difference (I–II)	Significance
$P_{pied\text{-}piped}$	*wh-*	*that*	0.845	$p < 0.001$
		Ø	1.043	$p < 0.001$
	that	*wh-*	−0.845	$p < 0.001$
		Ø	0.198	ns
	Ø	*Wh-*	−1.043	$p < 0.001$
		that	−0.198	ns
$P_{stranded}$	*wh-*	*that*	−0.092	ns
		Ø	−0.112	ns
	that	*wh-*	0.092	ns
		Ø	−0.020	ns
	Ø	*wh-*	0.112	ns
		that	0.020	ns

Rows with significant effects are shaded in grey.

- PREPOSITION PLACEMENT*PP TYPE.
  $F_1(2,70) = 14.213$, $p < 0.001$, $\eta^2 = 0.04$
  $F_2(2,10) = 6.344$, $p < 0.05$, $\eta^2 = 0.10$

While the results of the repeated measures ANOVA thus identified the same factors as influential in the Kenyan English data as in the British English study, a closer inspection of the two significant interactions by a post-hoc Tukey test helped to detect an interesting difference between the two varieties (see Tables 5.3 and 5.4).

On the one hand, Tables 5.3 and 5.4 contain results which match those of the British English study: pied-piping with *wh*-relativizers is significantly better than with *that* and Ø, but preposition-stranding is equally acceptable with all three relativizers. In addition to that, stranding with prepositional verbs is judged better than with locational/temporal and manner/frequency adjunct PPs.

On the other hand, unlike in the British English study, preposition-stranding with locational/temporal adjunct PPs does not receive significantly better scores than with manner/frequency adjunct PPs. Furthermore, pied-piping with the latter type of PP is judged better than with either locational/temporal adjunct PPs or prepositional verbs. In order to further investigate these effects, just like in Figures 5.1 and 5.2, mean judgements with standard error bars were plotted for pied-piping and stranding across RELATIVIZERS and PP TYPES (see Figures 5.5 and 5.6).

Table 5.4 *Result of post-hoc Tukey test for* PREPOSITION PLACEMENT**PP* TYPE *interaction*

Dependent variable	(I) PP type	(II) PP type	Mean difference	Significance
$P_{pied\text{-}pied}$	$V_{prepositional}$	$Adjunct_{loc/temp}$	−0.054	ns
		$Adjunct_{man/frequ}$	−0.406	$p < 0.005$
	$Adjunct_{loc/temp}$	$V_{prepositional}$	0.054	ns
		$Adjunct_{man/frequ}$	−0.352	$p < 0.005$
	$Adjunct_{man/frequ}$	$V_{prepositional}$	0.406	$p < 0.005$
		$Adjunct_{loc/temp}$	0.352	$p < 0.005$
$P_{stranded}$	$V_{prepositional}$	$Adjunct_{loc/temp}$	0.482	$p < 0.005$
		$Adjunct_{man/frequ}$	0.473	$p < 0.005$
	$Adjunct_{loc/temp}$	$V_{prepositional}$	−0.482	$p < 0.005$
		$Adjunct_{man/frequ}$	−0.009	ns
	$Adjunct_{man/frequ}$	$V_{prepositional}$	−0.473	$p < 0.005$
		$Adjunct_{loc/temp}$	0.009	ns

Rows with significant effects are shaded in grey.

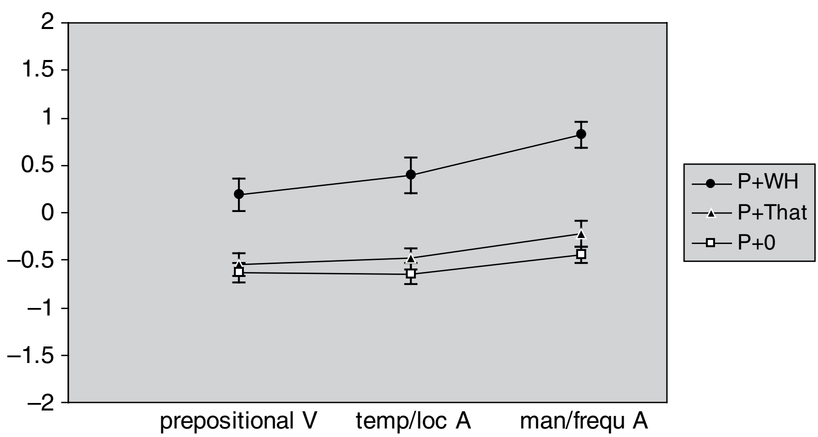

Figure 5.5 *Pied-piping across* RELATIVIZERS *and PP* TYPES *(Kenyan English speakers)*

Figure 5.5 shows that there is a steady increase in judgement scores for pied-piping with prepositional verbs at the lower end and manner/frequency adjunct PPs at the upper end of the scale (a trend that can be observed across relativizers). In contrast to this, Figure 5.6 indicates that stranding with

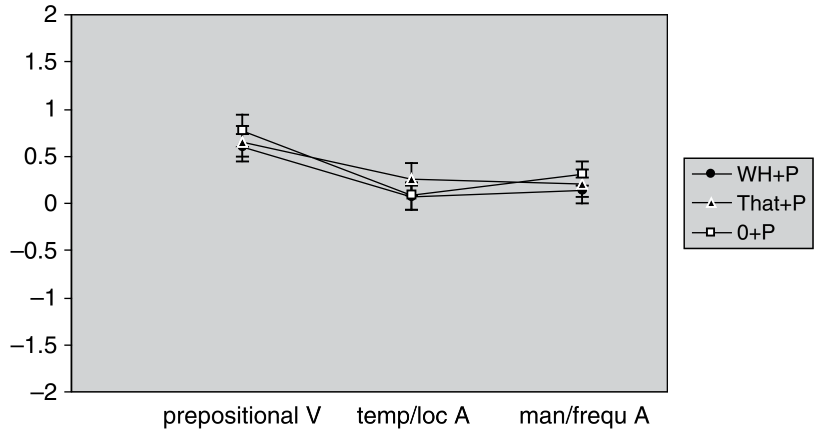

Figure 5.6 *Stranding across* RELATIVIZERS *and PP* TYPES *(Kenyan English speakers)*

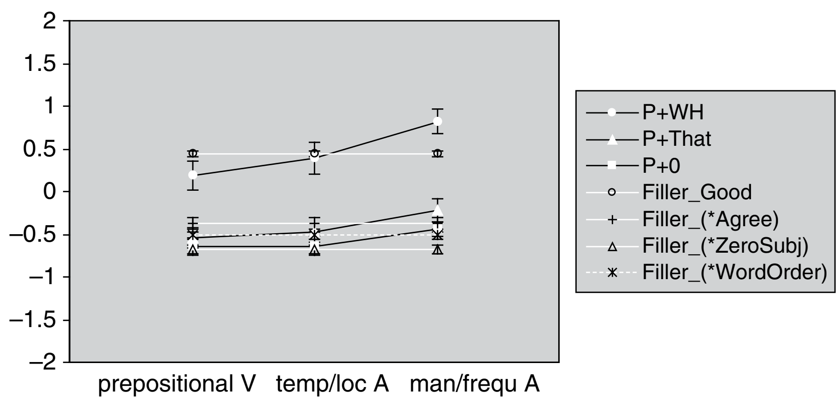

Figure 5.7 *Pied-piping across* RELATIVIZERS *and PP* TYPES *compared with fillers (Kenyan English speakers)*

locational/temporal adjunct PPs is judged as good as with manner/ frequency adjunct PPs.

Yet, how should these be results be interpreted? Are stranded prepositions with both locational/temporal adjunct and manner/frequency adjunct PPs grammatical in Kenyan English? In light of the results from the corpus study, this seemed extremely unlikely (since there the two types of PPs exhibited a markedly different effect; cf. sections 4.3.1.4 and 4.4.2). Again, it was necessary to compare the above results with the judgements of the grammatical and ungrammatical fillers to interpret them correctly (see Figure 5.7).

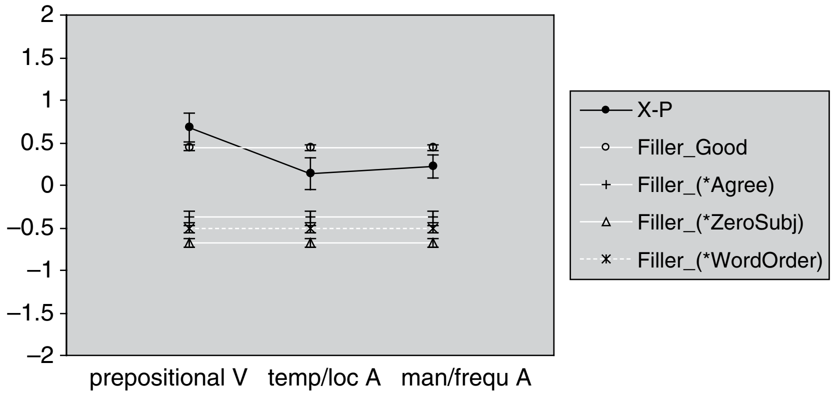

Figure 5.8 *Stranding means for all RELATIVIZERS across PP TYPES compared with fillers (Kenyan English speakers)*

Just like Figure 5.5, the mean judgements together with standard error bars of pied-piping across RELATIVIZERS and PP TYPES are plotted in Figure 5.7, but in addition to this the mean judgements for the various types of filler sentences are also given. Note first of all, that, just like in British English, pied-piping with *that* and *Ø* also appears to be a hard grammatical constraint violation in Kenyan English: the scores for both these relativizers pattern with the set of ungrammatical fillers. In contrast to British English, however, pied-piping with *wh*-relativizers is only as good as or better than the grammatical fillers with locational/temporal adjunct and manner/frequency adjunct PPs. In British English, pied-piping with prepositional verbs in *wh*-relative clauses had also received higher scores (cf. Figure 5.3). In the Kenyan data, on the other hand, prepositional verbs have judgement scores lower than the grammatical fillers (though the difference turns out to be non-significant: $t(34) = -1.258$, $p > 0.217$).[2]

The inclusion of the filler items then also helps in interpreting the effect of stranding with locational/temporal adjunct and manner/frequency adjunct PPs. In Figure 5.8 stranding with locational/temporal adjunct and manner/frequency adjunct PPs can be seen to be better than the set of ungrammatical fillers. Moreover, however, both PP type contexts also have scores lower than the one for the grammatical fillers. As a series of *t*-tests proves, both types of PPs are judged significantly worse than the grammatical fillers (locational/temporal adjunct PP – grammatical fillers: $t(34) = -2.924$, $p < 0.008$; manner/frequency adjunct PP – grammatical fillers: $t(34) = -2.894$, $p < 0.008$).

[2] All in all six *t*-tests were carried out on the data set: three tests comparing the scores for pied-piped *wh*-relative clauses across PP types against the set of grammatical fillers (cf. Figure 5.7), and three comparing the three PP types in Figure 5.8. The Bonferroni-corrected *p*-level for these *t*-tests was thus $0.05/6 = 0.0083$.

The results from the experiment thus corroborate the findings of the corpus study: pied-piping and stranding are both more entrenched in British English, which in this variety leads to higher scores of acceptability of pied-piping with prepositional verbs (which are lexically stored V-P sequences whose canonical word order is 'distorted' by pied-piping) and stranding with locational/temporal adjunct PPs (for which pied-piping would involve less processing cost). The Kenyan English results, on the other hand, indicate that the second-language variety favours the prototypical association of stranding with prepositional verbs and pied-piping with adjunct PPs, which both require less processing effort. In contrast to these processing-related soft constraints, the Kenyan data also identifies pied-piping with *that*- and *Ø*- as a hard grammatical constraint also operative in this variety.

Finally, however, it needs to be pointed out that, as can be expected from second-language data, the results from this experiment exhibit a greater noise component (a.k.a heterogeneity) than the British English one: the interactions PREPOSITION PLACEMENT*RELATIVIZER and PREPOSITION PLACEMENT*X-PP-RELATIONSHIP alone account for only 38% of the total by-subjects variability of the experiment (PREPOSITION PLACEMENT*RELATIVIZER: $\eta^2 = 0.25$ + PREPOSITION PLACEMENT*X-PP-RELATIONSHIP: $\eta^2 = 0.13$; cf. the 61 percent of variation accounted for in the British English data). In the by-item analysis the combined η^2-value of the two interactions explains a much greater percentage of the total variability (i.e. 72% = PREPOSITION PLACEMENT*RELATIVIZER: $\eta^2 = 0.47$ + PREPOSITION PLACEMENT*X-PP-RELATIONSHIP: $\eta^2 = 0.25$).

5.2 Preposition placement in relative clauses of varying complexity

In the next experiment it was decided to compare the judgements of pied-piping with *that*- and *Ø*-relativizers, on the one hand, with those with the *wh*-words *who* and *whom*, on the other. For the factor RELATIVIZER this meant testing the effect of four levels (*that* / *who* / *whom* / *Ø*). Besides this, the acceptability of doubled-preposition structures was investigated, leading to a three-level factor PREPOSITION PLACEMENT (stranded / pied-piped / doubled). Finally, it was decided to examine the influence of syntactic COMPLEXITY on preposition placement. In order to guarantee a clear-cut objective measurement of this factor, the following three levels of syntactic embedding were differentiated:

(5.5) **Levels of syntactic complexity**:
  a. I saw the teacher that Jane relied on.
  b. I saw the teacher that **[you claimed]** Jane relied on.
  c. I saw the teacher that **[you claimed that]** Jane relied on.

Table 5.5 *Sample token set*

Factor combination	Stimulus
Simple_Stranded_THAT_P	I saw the teacher that Jane relied on.
Simple_Stranded_WHO_P	I saw the teacher who Jane relied on.
Simple_Stranded_WHOM_P	I saw the teacher whom Jane relied on.
Simple_Stranded_ZERO_P	I saw the teacher Jane relied on.
Simple_P_THAT	I saw the teacher on that Jane relied.
Simple_P_WHO	I saw the teacher on who Jane relied.
Simple_P_WHOM	I saw the teacher on whom Jane relied.
Simple_P_ZERO	I saw the teacher on Jane relied.
Simple_P_THAT_P	I saw the teacher on that Jane relied on.
Simple_P_WHO_P	I saw the teacher on who Jane relied on.
Simple_P_WHOM_P	I saw the teacher on whom Jane relied on.
Simple_P_ZERO_P	I saw the teacher on Jane relied on.
Complex_THAT_That_P	I saw the teacher that you claimed that Jane relied on.
Complex_WHO_That_P	I saw the teacher who you claimed that Jane relied on.
Complex_WHOM_That_P	I saw the teacher whom you claimed that Jane relied on.
Complex_ZERO_That_P	I saw the teacher you claimed that Jane relied on.
Complex_P_THAT_That	I saw the teacher on that you claimed that Jane relied.
Complex_P_WHO_That	I saw the teacher on who you claimed that Jane relied.
Complex_P_WHOM_That	I saw the teacher on whom you claimed that Jane relied.
Complex_P_ZERO_That	I saw the teacher on you claimed that Jane relied.
Complex_P_THAT_That_P	I saw the teacher on that you claimed that Jane relied on.
Complex_P_WHO_That_P	I saw the teacher on who you claimed that Jane relied on.
Complex_P_WHOM_That_P	I saw the teacher on whom you claimed that Jane relied on.
Complex_P_ZERO_That_P	I saw the teacher on you claimed that Jane relied on.
Complex_THAT_Zero_P	I saw the teacher that you claimed Jane relied on.
Complex_WHO_Zero_P	I saw the teacher who you claimed Jane relied on.
Complex_WHOM_Zero_P	I saw the teacher whom you claimed Jane relied on.
Complex_ZERO_Zero_P	I saw the teacher you claimed Jane relied on.
Complex_P_THAT_Zero	I saw the teacher on that you claimed Jane relied.
Complex_P_WHO_Zero	I saw the teacher on who you claimed Jane relied.
Complex_P_WHOM_Zero	I saw the teacher on whom you claimed Jane relied.
Complex_P_ZERO_Zero	I saw the teacher on you claimed Jane relied.
Complex_P_THAT_Zero_P	I saw the teacher on that you claimed Jane relied on.
Complex_P_WHO_Zero_P	I saw the teacher on who you claimed Jane relied on.
Complex_P_WHOM_Zero_P	I saw the teacher on whom you claimed Jane relied on.
Complex_P_ZERO_Zero_P	I saw the teacher on you claimed Jane relied on.

In the experiment simple relative clauses (such as (5.5a)) were thus contrasted with syntactically more complex ones in which the relative clause is embedded within another clause. This clause was constructed either without an overt complementizer (as in (5.5b)) or with an overt *that*-subordinator (5.5c). Accordingly the three levels of the factor COMPLEXITY were named ‘simple’, ‘complex_zero’ and ‘complex_that’ (and are characterized by an increasingly complex bridge structure; see Gries 2002: 231 and section 3.5.).

This design resulted in PREPOSITION PLACEMENT × RELATIVIZER × COMPLEXITY = 3 × 4 × 3 = 36 cells. Since prepositional verbs had seemed to display the most interesting complexity effect (i.e. a higher percentage of stranding in more complex contexts) in the ICE corpus studies (cf. section 4.4.3), it was decided to focus on this PP type in the experiment in question, (a second reason being that most adjunct PPs would not have licensed a *who*(*m*)-relativizer; cf. **the place in whom I was*, **the way in who I did it*, etc.). In order to counterbalance the required thirty-six material sets, the following list of thirty-six prepositional verbs was taken from the *Longman Grammar of Spoken and Written English* (Biber *et al.* 1999: 403–23) and the *Comprehensive Grammar of the English Language* (Quirk *et al.* 1985: 1150–68):

(5.6) **Prepositional verbs**:
apologise for, ask for, belong to, call on, care for, collide with, depend on, disapprove of, dream of, focus on, hear about, hear of, laugh at, listen to, live with, look after, look at, look for, meet with, pick on, play with, point at, read of, refer to, rely on, shout at, sleep with, smile at, speak to, talk about, think about, think of, work for, worry about, write about, write to

Closely modelled on McDaniel and Cowart's experimental study on resumptive pronouns in English (1999),[3] the token sets for the experiment were created maximally simple: just as in McDaniel and Cowart's study, the stimuli were introduced by a short subject–verb sequence (*I saw … / That's … / Here's … / There's … / There goes …*) which was followed by a simple object NP (*the teacher*, *the man*, *the singer*, etc.). The relative clause containing the prepositional verb then modified this NP. Table 5.5 illustrates how this sentence structure together with the required factor combinations created all the members of token set for the verb *rely on*. As the table shows, with *rely on* the bridge structure *you claimed* (*that*) was used to test the effect of the factor COMPLEXITY. Just as with the prepositional verbs, the following thirty-six different structures were employed for the other stimuli to preclude idiosyncratic lexical effects:[4]

(5.7) **Bridge structures**:
you claimed, I imagined, I read, I observed, you feared, I doubted, I didn't know, I explained, I told you, I thought, you said, I suggested, you admitted, you knew, you felt, you guessed, I swore, I remembered, I heard, you insisted, you believed, you reported, I denied, I noted, you remarked, you revealed, I mentioned, I declared, you implied, you discovered, I suspected, you expected, I hoped, you expected[5], I learnt, you pointed out

[3] I am very grateful to Wayne Cowart for sharing his experimental stimuli with me, which greatly facilitated the design of this experiment.

[4] Note that due to the design of the experiment potential idiosyncratic effects of different bridge structures (cf. section 3.5.1) could not be investigated: since individual bridge structures always co-occurred with only one prepositional verb, the independent effect of the two factors cannot be assessed. However, as a result of the counterbalanced stimuli set (see section 2.3.1), it could at least be ensured that any such idiosyncratic effect was spread over all conditions and thus did not affect the results of the experiment.

[5] This item was the only bridge structure that was used twice in the experiment.

Table 5.6 *Filler set A*

Factor combination	Filler stimulus
Simple_That_S	I saw the drunk that scared Fran.
Simple_Who_S	I saw the professor who angered Nina.
Simple_Whom_S	There's the saleswoman whom married Steve.
Simple_Zero_S	There's the waiter insulted Jacqueline.
Simple_That_O	I saw the nurse that the child hit.
Simple_Who_O	I saw the woman who John envied.
Simple_Whom_O	That's the principal whom Lisa interviewed.
Simple_Zero_O	There is the policewoman the thief found.
Complex_That_That_S	That's the teacher that I noticed that annoyed Martin.
Complex_Who_That_S	That's the actor who I understood that dated Gina.
Complex_Whom_That_S	There is the patient whom you guessed that called the dentist.
Complex_Zero_That_S	That's the taxi driver you remembered that located Mike.
Complex_That_Zero_S	There's the actress that you knew encountered Mark.
Complex_Who_Zero_S	That's the girl who you questioned called Fred.
Complex_Whom_Zero_S	Here's the child whom I didn't comprehend bothered Sam.
Complex_Zero_Zero_S	That's the man you said rescued the teenager.
Complex_That_That_O	That's the child that I considered that Joe liked.
Complex_Who_That_O	There's the captain who I told you that Ann dated.
Complex_Whom_That_O	That's the secretary whom I challenged that Bill harassed.
Complex_Zero_That_O	There goes the little boy I didn't know that the teacher hugged.
Complex_That_Zero_O	There goes the boy that I wondered Sue punched.
Complex_Who_Zero_O	There is the man who you explained the doctor requested.
Complex_Whom_Zero_O	That's the woman whom I heard the baby kissed.
Complex_Zero_Zero_O	There goes the reporter I read Janet sued.

Once token sets like Table 5.5 had been created for all prepositional verbs, the stimuli were counterbalanced, which resulted in thirty-six material sets containing all factor combinations but never with the same lexicalization.

Due to the fact that the stimuli conspicuously differed in complexity, it was decided to create the set of fillers in a similar fashion so as to preclude subjects guessing the aim of the experiment and forming confounding implicit hypotheses. For this, relative clauses with relativized subject- and object-positions were also crossed with the factors RELATIVIZER and COMPLEXITY to give a first set of twenty-four fillers (Table 5.6). The procedure was repeated to get a complete set of forty-eight fillers (and thus a filler:stimuli ratio of 1.5:1; cf. Appendix A.5.2.1 for a sample material set including all fillers). As can be seen in Table 5.6, this filler set contained perfectly grammatical sentences (e.g. all relative clauses with an object gap 'O') as well as ungrammatical ones (case violations such as *whom* in subject position, e.g. *There's the saleswoman whom married Steve* or subject contact clauses, e.g. *There's the waiter insulted Jacqueline*).

As with all studies, the experiment employed the Magnitude Estimation method, and all experimental data were normalized by transformation to *z*-scores and subjected to repeated measures analyses of variance. Since all in all eighty-four stimuli (thirty-six experimentals plus forty-eight fillers) had to be judged in addition to the usual training sets, the experiment turned out to be rather long, lasting thirty to forty-five minutes. Due to this (and the fact that no financial compensation for their time could be offered to subjects), it was only possible to get thirty-six British and Kenyan speakers each to take part. The sample size for both varieties thus met the required minimum for statistical analysis via the parametric repeated measures ANOVA ($n > 30$). Moreover, this ensured that each material set was judged by an informant (cf. Cowart 1997: 82). It precluded the possibility of a separate by-item and by-subject analysis though, since each material set was judged by only one person. For this reason, the results for this experiment will only give a single F-value, which has to be interpreted as a combined effect of item- and subject-related factors. Finally, the experiment was again designed using the WebExp software, but due to the low response rate to the online version, a printed questionnaire was administered to both British and Kenyan speakers. As in the preceding section, I will first give an overview of the results for the former group of speakers, and then present those of the Kenyan ones.

5.2.1 British English

Thirty-six undergraduate students and lecturers from the universities of Edinburgh and Central Lancashire, all of them native speakers of British English, were recruited for the first run of the experiment. These comprised twenty-eight female and eight male informants, displaying an age range of 19–65. As the repeated measures ANOVA showed, these between-subject factors were non-significant: AGE at $F(1,33) = 0.843$, $p > 0.35$ and GENDER at $F(1,33) = 0.275$, $p > 0.60$. In contrast to this, the set of within-subject factors yielded the following significant results (see Appendix A.5.2.2 for the full analysis):

- COMPLEXITY:
  $F(2,70) = 20.993$, $p < 0.001$, $\eta^2 = 0.02$
- PREPOSITION PLACEMENT:
  $F(1.545,54.084) = 154.640$, $p < 0.001$, $\eta^2 = 0.29$
- RELATIVIZER:
  $F(2.485,86.961) = 32.066$, $p < 0.001$, $\eta^2 = 0.07$
- COMPLEXITY*PREPOSITION PLACEMENT:
  $F(4,140) = 5.488$, $p < 0.001$, $\eta^2 = 0.01$
- PREPOSITION PLACEMENT*RELATIVIZER:
  $F(6,210) = 21.550$, $p < 0.001$, $\eta^2 = 0.06$
- COMPLEXITY*PREPOSITION PLACEMENT*RELATIVIZER.
  $F(11.108,388.773) = 3.092$, $p< 0.005$, $\eta^2 = 0.01$

In order to interpret these results, in a first step post-hoc Tukey tests were carried out for the two-way interactions. Table 5.7 presents the results for the Tukey test of the COMPLEXITY*PREPOSITION PLACEMENT interaction. As this table illustrates, the factor COMPLEXITY has different effects depending on

Table 5.7 *Result of post-hoc Tukey test for* COMPLEXITY*PREPOSITION PLACEMENT *interaction*

Dependent variable	(I) Relativizer	(II) Relativizer	mean difference	Significance
$P_{stranded}$	simple	complex_C_*that*	0.637	$p < 0.001$
		complex_C_Ø	0.446	$p < 0.001$
	complex_C_*that*	simple	−0.637	$p < 0.001$
		complex_C_Ø	−0.191	ns
	complex_C_Ø	simple	−0.446	$p < 0.001$
		complex_C_*that*	0.191	ns
P_{piped}	simple	complex_C_*that*	0.240	$p < 0.05$
		complex_C_Ø	0.175	ns
	complex_C_*that*	simple	−0.240	$p < 0.05$
		complex_C_Ø	−0.065	ns
	complex_C_Ø	simple	−0.175	ns
		complex_C_*that*	0.065	ns
$P_{doubled}$	simple	complex_C_*that*	0.163	ns
		complex_C_Ø	0.154	ns
	complex_C_*that*	simple	−0.163	ns
		complex_C_Ø	−0.009	ns
	complex_C_Ø	simple	−0.154	ns
		complex_C_*that*	0.009	ns

Rows with significant effects are shaded in grey.

PREPOSITION PLACEMENT: doubled prepositions receive equal scores regardless of the level of syntactic complexity. Next, stranded prepositions receive significantly higher scores in simple relative clauses than in more complex ones. It makes no difference, however, whether the bridge structure contains an overt complementizer (complex_C_*that*: e.g., *you claimed that*) or not (complex_C_*Ø*: e.g. *you claimed*). Pied-piped prepositions, on the other hand, are also judged best in simple relative clauses, but only the presence of a *that*-complementizer leads to significantly lower scores. These effects are best explained by taking a look at the examples in (5.8) and (5.9):

(5.8) 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**.

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

In both (5.8b,c) and (5.9b,c) the presence of a bridge structure makes the sentences more difficult to parse than the simple ones in (5.8a/5.9a). Remember, however, that the pied-piped versions in (5.9) have a clause-initial preposition that is closely lexically associated with the main verb. On the one hand, pied-piping this preposition avoids potential garden-paths effects (there is no attempt to integrate *on whom* as the filler of *claimed* since pied-piping obeys Hawkins's 2004 Avoid Competing Subcategorizor principle; cf. section 3.5). In (5.9c), however, the intervening *that*-complementizer is one additional syntactic item that defers the integration of *on* and *rely* as a single lexical item, which accounts for the lower acceptability of these constructions compared to a bridge structure without an overt complementizer.

Next, Table 5.8 provides an overview of the Tukey test results for the interaction of the factors PREPOSITION PLACEMENT*RELATIVIZER. For the stranded data, Table 5.8 confirms the results of the first experiment: preposition-stranding is equally acceptable with all relativizers. In contrast to this, as expected, pied-piping with *whom* is judged better than with all other relativizers. In addition to this, however, Table 5.8 also indicates that pied-piping with *who* is significantly better than with *that* or *Ø* (which receive statistically equal scores). It thus seems that the ungrammaticality of **the teacher on that Jane relied* is not the result of case mismatch on a par with **the teacher on who Jane relied* (as argued by Payne and Huddleston 2002 or Sag 1997; cf. section 3.1.2.1), but a word order violation similar to **the teacher on Ø Jane relied*. This is corroborated by the results for the ungrammatical doubled-preposition structures: these are judged better with *who* and *whom* (e.g. **the teacher on who(m) Jane relied on*) than with *that* or *Ø* (cf. **the teacher on (that) Jane relied on*).

For the further investigation of these effects, it was again decided to plot the results of the various factor interactions, giving the mean judgements of conditions together with standard error bars. Since the set of fillers always plays an important role in the interpretation of effects, these will be looked at first. I will focus on fillers with relativized subject positions, which contain grammatical (with *that* and *who* relativizers) as well as ungrammatical stimuli (case mismatch effects with *whom* in subject position and subject contact clauses).

Figure 5.9 gives the results for the different relativizers in subject position of simple relative clauses in filler sets 1 and 2. While the *that*-, *who*- and *whom*-stimuli were judged equally in both filler sets, the subject contact clause in filler set 1 (*I saw the nurse Ø hit the child*) received much better scores than the one in filler set 2 (*There's the waiter Ø insulted Jacqueline*). The reason for this was apparently that the sentence of filler set 1 had an alternative non-relative clause reading ('I saw how the nurse hit the child'), which subjects preferred. For the present analysis, filler set 2 is therefore the better reference for simple subject relative clauses.

Table 5.8 *Result of post-hoc Tukey test for* PREPOSITION PLACEMENT*RELATIVIZER *interaction*

Dependent variable	(I) Complex	(II) Complex	Mean difference	Significance
$P_{stranded}$	that	who	0.117	ns
		whom	0.203	ns
		Ø	−0.009	ns
	who	that	−0.117	ns
		whom	0.086	ns
		Ø	−0.126	ns
	whom	that	−0.203	ns
		who	−0.086	ns
		Ø	−0.212	ns
	Ø	that	0.009	ns
		who	0.126	ns
		whom	0.212	ns
P_{piped}	that	who	−0.753	$p < 0.001$
		whom	−1.008	$p < 0.001$
		Ø	0.160	ns
	who	that	0.753	$p < 0.001$
		whom	−0.329	$p < 0.05$
		Ø	0.913	$p < 0.001$
	whom	that	1.008	$p < 0.001$
		who	0.329	$p < 0.05$
		Ø	1.024	$p < 0.001$
	Ø	that	−0.160	ns
		who	−0.913	$p < 0.001$
		whom	−1.024	$p < 0.001$
$P_{doubled}$	that	who	−0.526	$p < 0.001$
		whom	−0.724	$p < 0.001$
		Ø	0.183	ns
	who	that	0.526	$p < 0.001$
		whom	−0.198	ns
		Ø	0.709	$p < 0.001$
	whom	that	0.724	$p < 0.001$
		who	0.198	ns
		Ø	0.908	$p < 0.001$
	Ø	that	−0.183	ns
		who	−0.709	$p < 0.001$
		whom	−0.908	$p < 0.001$

Rows with significant effects are shaded in grey.

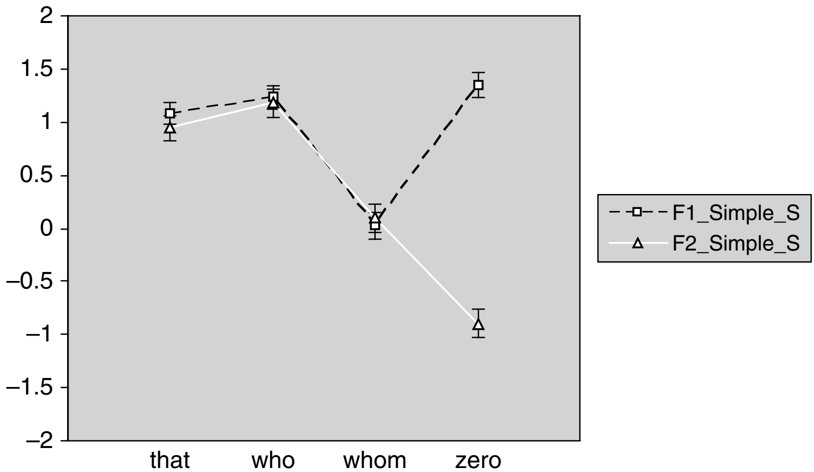

Figure 5.9 *Means for fillers with relativized subject position in simple relative clauses across filler sets 1 and 2 (British English data)*

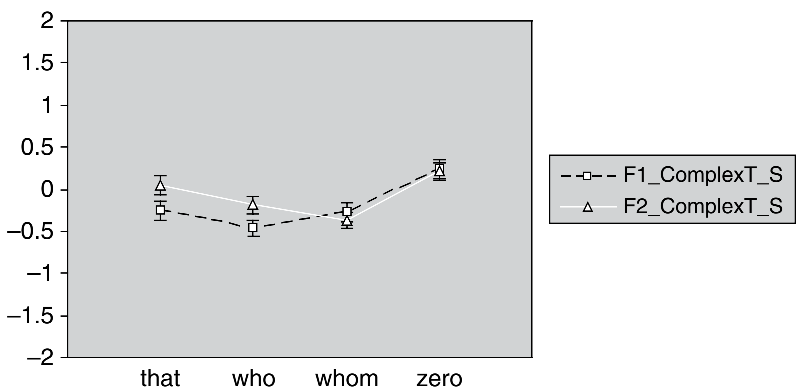

Figure 5.10 *Means for fillers with relativized subject position in complex relative clauses with a* that-*complementizer across filler sets 1 and 2 (British English data)*

In contrast to this, as Figure 5.10 shows, both filler sets yield comparable results for relative clauses with relativized subject position embedded in a bridge structure with a *that*-complementizer (e.g. *There is the patient* __ [*you guessed that*] *called the dentist*).

Finally, the two filler sets again give divergent results for relative clauses with relativized subject position embedded in a bridge structure with a Ø-complementizer (e.g. *There is the patient* __ [*you guessed*] *called the dentist*). The most notable difference between the filler sets in Figure 5.11 concerns the *whom*- and Ø-stimuli. The stimuli in question are given in (5.10) and (5.11):

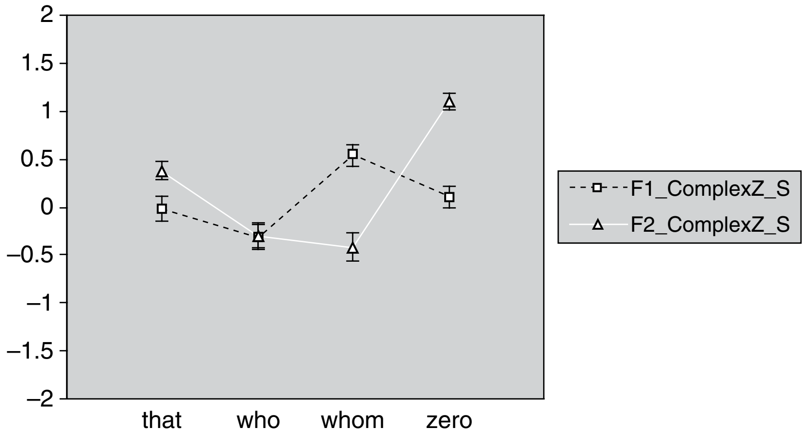

Figure 5.11 *Means for fillers with relativized subject position in complex relative clauses with a Ø-complementizer across filler sets 1 and 2 (British English data)*

(5.10) *whom*-subject in relative clause with Ø-complementizer bridge structure:
 a. Filler set 1: That's the man whom you said Ø rescued the teenager.
 b. Filler set 2: Here's the child whom I didn't comprehend Ø bothered Sam.

(5.11) Ø-subject in relative clause with Ø-complementizer bridge structure:
 a. Filler set 1: That's the child Ø I considered Ø liked Joe.
 b. Filler set 2: That's the man Ø you said Ø rescued the teenager.

Comparing (5.10) and (5.11) with Figure 5.11, it becomes apparent that it is the prototypical bridge verb *say* that leads to higher scores (for the *whom*-subjects in filler set 1 and the Ø-subjects in filler set 2). The alternative sentences always employ less typical bridge verbs (*comprehend* and *consider*). This might indicate an interesting field for further research; however, for the present study it was decided to use the more prototypical instances with *say* (i.e. (5.10a) and (5.10b)) as reference points.

In light of these findings, Figure 5.12 gives a representative overview of fillers with relativized subject positions (using filler set 2 for the simple and complex-*that* contexts as well as the Ø-subject example in relative clauses with Ø-complementizer (i.e. (5.10b)), and filler set 1 for all other Ø-complementizer structures). Figure 5.12 first of all identifies simple subject contact clauses (*There's the waiter Ø insulted Jacqueline*) as the least acceptable structure. In addition to this, as suggested by Huddleston, Pullum and Peterson (2002: 1047), in embedded structures with Ø-complementizers Ø-subjects become grammatical (cf. *That's the man Ø you said Ø rescued the teenager*). A similar, although not as considerable, increase in acceptability appears in subject clauses with *whom* (which also conforms to Huddleston, Pullum and Peterson's predictions (2002: 1047)). Note, however, that for both phenomena the presence of a *that*-complementizer reduces this effect

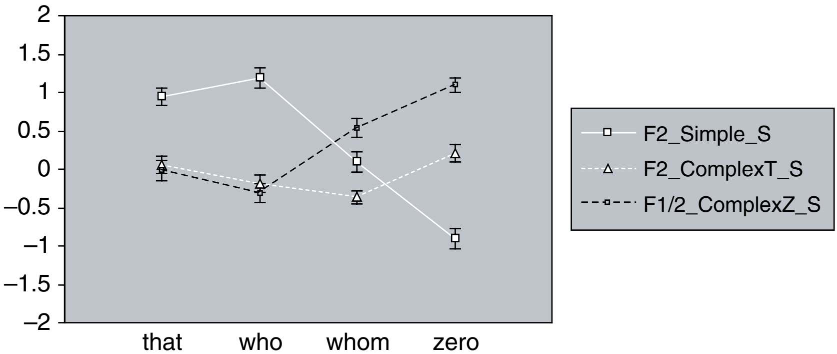

Figure 5.12 *Means for fillers with relativized subject position across* COMPLEXITY *for selected fillers (British English data)*

(in the case of *whom* even leading to lower acceptability scores). Finally, for the grammatical *that*- and *who*- subjects, Figure 5.12 illustrates how an increase in complexity leads to lower judgement scores, but never as low as the ungrammatical simple subject contact clauses. All of these results for the filler stimuli probably warrant an experimental investigation of their own; for the present study, however, it is especially the effect of simple subject-relative clauses with Ø- and *whom*-subjects that will be used as a reference point: while both can be considered hard grammatical constraint violations, their different results allow comparison of structural violations such as the former (i.e. the lack of an obligatory element) and case mismatch effects such as the latter (i.e. the use of an oblique form instead of a nominative one).

Returning to the results of the experiment, Figure 5.13 plots the scores for PREPOSITION PLACEMENT across RELATIVIZERS in simple relative clauses. As can be seen in, only pied-piping with *whom* gives results as good as those for stranding (since pied-piping with *whom* = stranding with *whom*, and stranding being statistically equal for all relativizers). In all other cases, the scores for pied-piped and doubled prepositions approximate each other. In addition to this, pied-piping with *who* is judged better than with *that* and Ø.

In order to interpret these results, the mean scores for the simple *whom*-subject relative clauses (using the means for filler sets 1 and 2, which had an identical effect) as well as simple Ø-subject relative clauses (just filler set 2; cf. above) were included in Figure 5.14 (which focuses on stranding and pied-piping). Figure 5.14 shows that pied-piping with *that* and Ø patterns with the simple subject contact clauses at the end of the acceptability scale. Pied-piping with *who*, on the other hand, is clearly better than this and, as expected, is similar to the case mismatch violation of *whom*-subjects. The reason why pied-piping with *who* gets slightly worse results than *whom*-subjects is probably due to the fact that the former additionally involve

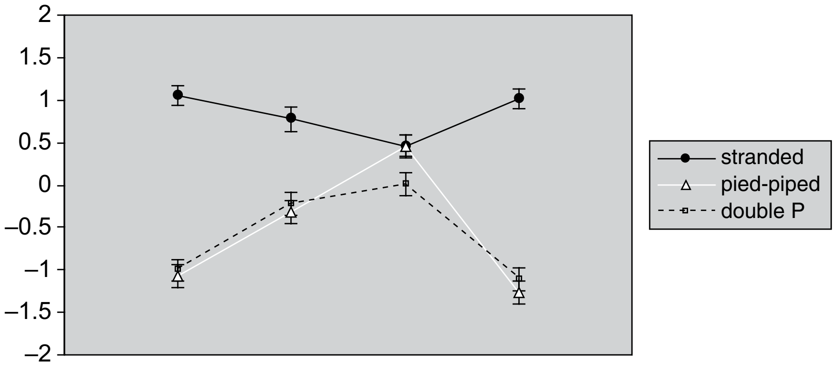

Figure 5.13 *Means for* PREPOSITION PLACEMENT *across* RELATIVIZERS *in simple relative clauses (British English data)*

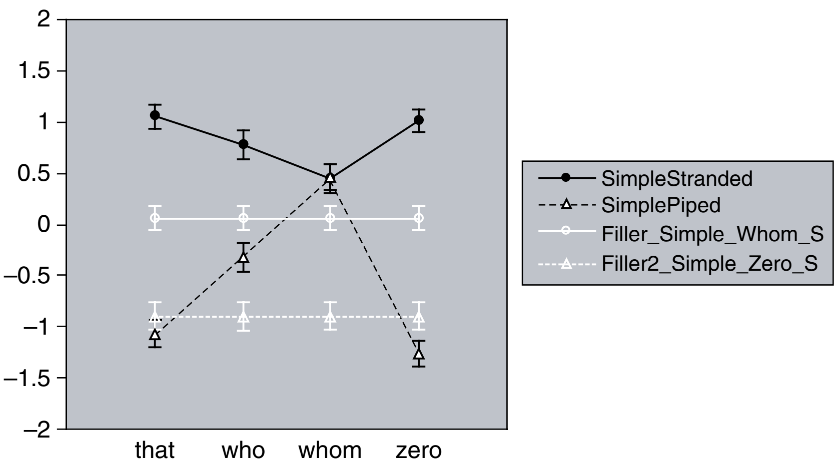

Figure 5.14 *Means for* PREPOSITION PLACEMENT *across* RELATIVIZERS *in simple relative clauses plus selected ungrammatical fillers (British English data)*

prepositional verbs (i.e. a preposition that is deferred from its lexically associated verb).

The effect of bridge verbs with *that*-complementizers is then visualized in Figure 5.15. Stranding with prepositional verbs is, as expected, also preferred in these complex clauses since this constitutes the easiest way of identifying a preposition with its lexically associated verb. As illustrated, this effect even prevails over the advantage of avoiding garden-path effects by pied-piping: in Figure 5.15 pied-piping with *whom* is clearly judged worse

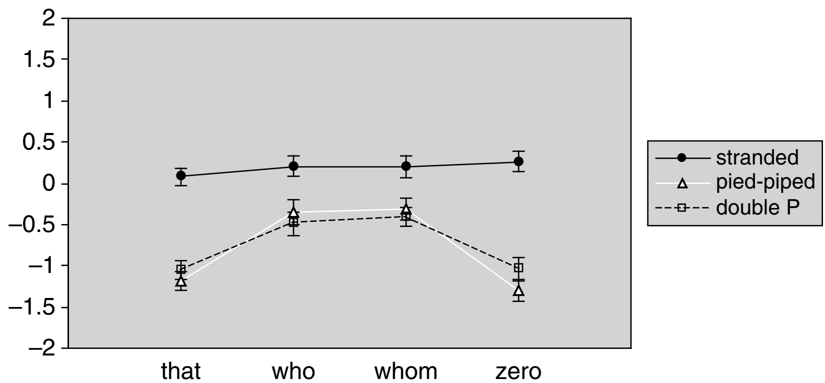

Figure 5.15 *Means for* PREPOSITION PLACEMENT *across* RELATIVIZERS *in complex* that-*complementizer relative clauses (British English data)*

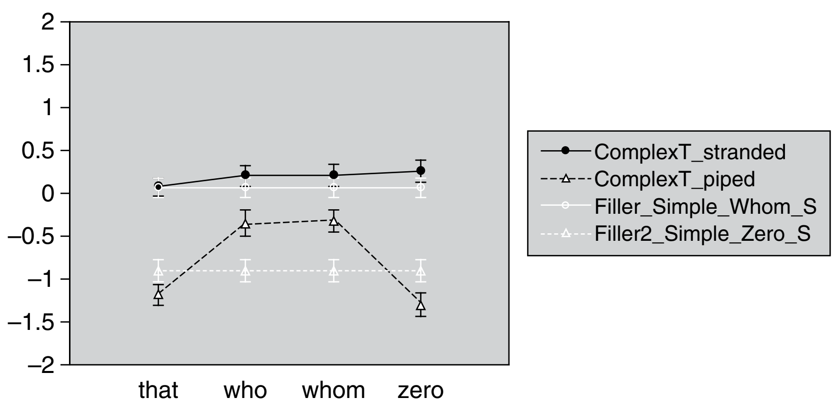

Figure 5.16 *Means for* PREPOSITION PLACEMENT *across* RELATIVIZERS *in complex* that-*complementizer relative clauses plus selected ungrammatical fillers (British English data)*

than the stranded alternative. Moreover, in this context pied-piping receives scores similar to preposition doubling.

Figure 5.16 compares these results for stranding and pied-piping with the chosen set of ungrammatical fillers. The increased complexity leads to stranding getting scores similar to simple *whom*-subject fillers (indicating that low scores in Magnitude Estimation experiments do not automatically entail ungrammaticality, but require careful examination of all factors involved). Pied-piping with *who* and *whom*, however, is still clearly better than with *that* and *Ø*, which again pattern with the ungrammatical subject contact clauses.

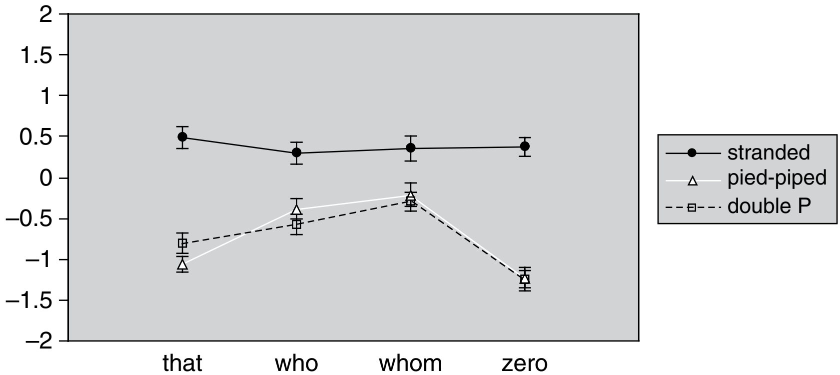

Figure 5.17 *Means for PREPOSITION PLACEMENT across RELATIVIZERS in complex Ø-complementizer relative clauses (British English data)*

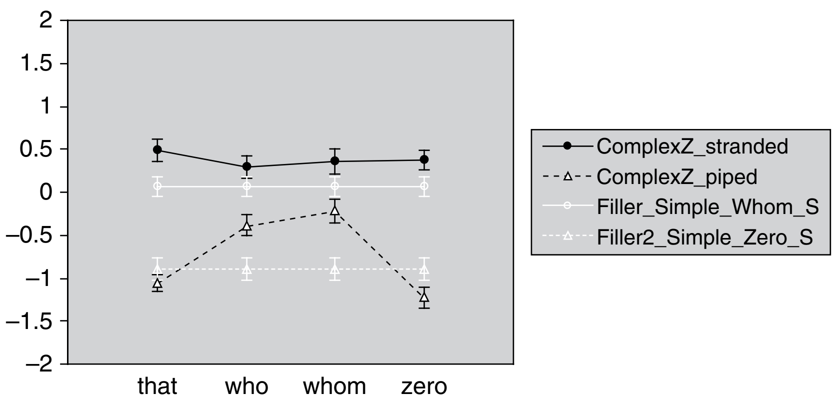

Figure 5.18 *Means for PREPOSITION PLACEMENT across RELATIVIZERS in complex Ø-complementizer relative clauses plus selected ungrammatical fillers (British English data)*

Finally, the effect of bridge verbs with Ø-complementizers (Figure 5.17) turns out to be similar to those with *that*-complementizers. In Figure 5.17, just like in Figure 5.15, the increased complexity leads to stranding being favoured over pied-piping even with *whom*. Besides this, doubled and pied-piped prepositions are judged equally.

The graph corresponding to Figure 5.16 is then given in Figure 5.18. As a comparison of figures 5.18 and 5.16 reveals, the absence of a *that*-complementizer leads to a slight increase of acceptability of the grammatical structures. The low judgement scores for pied-piping with *that* and Ø, however, are unaffected by this.

Table 5.9 *T-test results P +* who/that-*RELATIVIZER vs simple Ø-subject fillers*

COMPLEXITY	P + X	*filler	Mean difference	*T*-test
simple	P + *who*	*simple Ø-subject	0.581	$t(35) = 2.940, p < 0.008$
	P + *that*	*simple Ø-subject	-0.168	$t(35) = 0.825, p > 0.40$
that-complex	P + *who*	*simple Ø-subject	0.550	$t(35) = 2.855, p < 0.008$
	P + *that*	*simple Ø-subject	-0.284	$t(35) = 1.712, p > 0.05$
Ø-complex	P + *who*	*simple Ø-subject	0.513	$t(35) = 2.831, p < 0.008$
	P + *that*	*simple Ø-subject	-0.162	$t(35) = 0.873, p > 0.35$

Bonferroni-adjusted p-level: $p = 0.05/6 = 0.0083$

This experiment has again corroborated the hypothesis that preposition-stranding is equally good across all relativizers in British English. In addition to this, syntactic complexity has the expected effect with prepositional verbs, namely that stranding is favoured over pied-piping since it facilitates the integration of lexicalized verb–preposition structures. Furthermore, the experiment again proved pied-piping with *that*- and Ø-relativizers to be hard grammatical constraint violations. Yet, in addition to this, the results for pied-piping with *who* have shown that the effects of *that* and Ø are not reducible to case mismatch violations. Instead, these are simply structural configurations which are not licensed by the grammar (in contrast to P + *who*, which can be generated by the grammar but results in a case mismatch). This claim can be further supported by comparing the most ungrammatical Ø-subject fillers with P + *who* and P + *that* structures across the factor COMPLEXITY in a series of *t*-tests (Table 5.9). The column 'Mean difference' shows that P + *who* always gets higher scores than the ungrammatical Ø-subject fillers, and that this effect is significant even at the Bonferroni-adjusted *p*-level of 0.0083. P + *that*, on the other hand, receives scores lower than the Ø-subject fillers, but the *t*-test indicates that this difference in means is not significant.

Note that the importance of the above two effects, i.e. stranding being generally preferred with prepositional verbs and the ban on P + *that*- and P+ Ø-structures, is illustrated by the fact that the best model for the data, which accounts for 67 percent of the overall variation, is the one only containing the factors PREPOSITION PLACEMENT ($\eta^2 = 0.55$) and PREPOSITION PLACEMENT*RELATIVIZER ($\eta^2 = 0.12$).

Finally, with the exception of simple *whom*-relative clauses, all relativizers had identical effects for pied-piping and preposition-doubling, despite the fact that the latter option is not considered grammatical. Note, however, that doubled-preposition structures have an interesting characteristic: while the construction as a whole is not part of the grammar of

Standard English (cf. **the person on whom she relied on*), the two individual instances of the preposition (the pied-piped part *on whom* and the stranded preposition *relied on*) are grammatical ('locally good' to extend an analysis of resumptive pronouns by Sam Featherston, p.c., to preposition placement). This explains why doubled prepositions yield similar results for all instances where the pied-piping part leads to ungrammaticality: in **the person on that she relied on* the sequence *on that* causes a severe structural violation and in **the person on who she relied on* the *on who* part violates case-matching principles. With *whom*-relativizers, on the other hand, both instances of the preposition in doubled structures are acceptable individually, and were probably not too conspicuous in complex clauses, resulting in similar scores for doubled and pied-piped prepositions in these contexts.

5.2.2 Kenyan English

The same experiment on complexity was administered to thirty-six speakers of Kenyan English (twenty-five female, eleven male; age range 20–26), all students at the University of Nairobi, in September 2006. These again only included eight speakers of a non-Bantu language (all native speakers of Dholou), which again precluded an in-depth test of potential first-language effects. The repeated measures ANOVA showed that the other between-subject factors AGE and GENDER were non-significant at $F(1,33) = 0.236$, $p > 0.60$ and $F(1,33) = 0.417$, $p > 0.50$, respectively. The significant within-subject factors were the following:

- COMPLEXITY:
  $F(2,70) = 11.410$, $p < 0.001$, $\eta^2 = 0.01$
- PREPOSITION PLACEMENT:
  $F(1.766,61.813) = 57.579$, $p < 0.001$, $\eta^2 = 0.12$
- RELATIVIZER:
  $F(3,105) = 17.820$, $p < 0.001$, $\eta^2 = 0.04$
- COMPLEXITY*PREPOSITION PLACEMENT:
  $F(4,140) = 3.439$, $p < 0.05$, $\eta^2 = 0.01$
- COMPLEXITY*RELATIVIZER:
  $F(6,210) = 3.529$, $p < 0.005$, $\eta^2 = 0.01$
- PREPOSITION PLACEMENT*RELATIVIZER:
  $F(6,210) = 9.357$, $p < 0.001$, $\eta^2 = 0.04$
- COMPLEXITY*PREPOSITION PLACEMENT*RELATIVIZER:
  $F(12,420) = 1.791$, $p < 0.05$, $\eta^2 = 0.01$

As in the British English study, the two interactions involving PREPOSITION PLACEMENT were subjected to post-hoc Tukey tests, yielding the results summarized in Tables 5.10 and 5.11. Just as in the British English data, the judgement of doubled prepositions is not affected by the factor COMPLEXITY.

Table 5.10 *Result of post-hoc Tukey test for* COMPLEXITY*PREPOSITION PLACEMENT *interaction*

Dependent variable	(I) Complex	(II) Complex	Mean difference	Significance
$P_{stranded}$	simple	complex_C_*that*	0.506	$p < 0.001$
		complex_C_Ø	0.201	ns
	complex_C_*that*	simple	−0.506	$p < 0.000$
		complex_C_Ø	−0.305	$p < 0.05$
	complex_C_Ø	simple	−0.201	ns
		complex_C_*that*	0.305	$p < 0.05$
P_{piped}	simple	complex_C_*that*	0.208	ns
		complex_C_Ø	0.170	ns
	complex_C_*that*	simple	−0.208	ns
		complex_C_Ø	−0.037	ns
	complex_C_Ø	simple	−0.170	ns
		complex_C_*that*	0.037	ns
$P_{doubled}$	simple	complex_C_*that*	0.018	ns
		complex_C_Ø	0.041	ns
	complex_C_*that*	simple	−0.018	ns
		complex_C_Ø	0.022	ns
	complex_C_Ø	simple	−0.041	ns
		complex_C_*that*	−0.022	ns

All rows with significant effects are shaded in grey.

Above it was argued that these structures combine two variants which both are individually locally acceptable and reduce online processing cost (i.e. a pied preposition, which avoids garden-path effects, and a stranded preposition, which facilitates the integration of the lexicalized verb–preposition structure). This account explains why processing effects do not affect the judgement scores of doubled prepositions.

In contrast to the British English results, however, complexity did not influence the Kenyan speakers' judgements of pied-piping. Moreover, with the stranded data, only a bridge structure with an overt *that*-complementizer led to a significantly lower acceptability score (compared to the British English data, in which both bridge structures caused lower scores). In order to interpret these effects, it will be necessary to compare them to the judgements of the ungrammatical fillers (see below).

Next, the post-hoc Tukey results for the interaction PREPOSITION PLACEMENT*RELATIVIZER are given in Table 5.11. Again, the results for the doubled-preposition structures pattern with those of the British English study: the scores for *whom* and *who* are equally high and both are better than

Table 5.11 *Result of post–hoc Tukey test for* PREPOSITION PLACEMENT*RELATIVIZER *interaction*

Dependent variable	(I) Relativizer	(II) Relativizer	Mean difference	Significance
$P_{stranded}$	that	who	0.289	ns
		whom	0.032	ns
		Ø	−0.148	ns
	who	that	−0.289	ns
		whom	−0.257	ns
		Ø	−0.437	$p < 0.01$
	whom	that	−0.032	ns
		who	0.257	ns
		Ø	−0.180	ns
	Ø	that	0.148	ns
		who	0.437	$p < 0.01$
		whom	0.180	ns
P_{piped}	that	who	−0.245	ns
		whom	−0.615	$p < 0.001$
		Ø	0.160	ns
	who	that	0.245	ns
		whom	−0.370	$p < 0.01$
		Ø	0.405	$p < 0.005$
	whom	that	0.615	$p < 0.001$
		who	0.370	$p < 0.01$
		Ø	0.775	$p < 0.001$
	Ø	that	−0.160	ns
		who	−0.405	$p < 0.005$
		whom	−0.775	$p < 0.001$
$P_{doubled}$	that	who	−0.569	$p < 0.001$
		whom	−0.809	$p < 0.001$
		Ø	−0.156	ns
	who	that	0.569	$p < 0.001$
		whom	−0.240	ns
		Ø	0.413	$p < 0.005$
	whom	that	0.809	$p < 0.001$
		who	0.240	ns
		Ø	0.653	$p < 0.001$
	Ø	that	0.156	ns
		who	−0.413	$p < 0.005$
		whom	−0.653	$p < 0.001$

Rows with significant effects are shaded in grey.

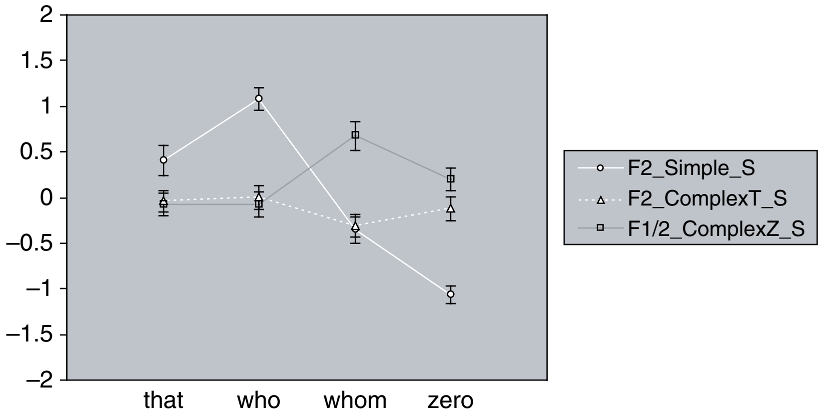

Figure 5.19 *Means for fillers with relativized subject position across* COMPLEXITY *for selected fillers (Kenyan English data)*

those for *that* and Ø. While *whom* is also preferred over the other relativizers in pied-piped relative clauses, the results for *who* differ from those of the British English study: pied-piping with *who* does not differ significantly from pied-piping with *that*. In addition to this, in stranded contexts Ø-relativizers are significantly better than *who*. Generally, this might be taken as an indication that *who*-relativizers might be deemed best avoided in Kenyan English, with speakers preferring Ø-relativizers in stranded and *whom* in pied-piped contexts. The precise nature of this effect will have to be assessed by comparing it to the set of ungrammatical fillers.

For the sake of comparison the same set of fillers were used as a reference point as in the British English study. Thus Figure 5.19, which plots the results of the means of judgements of fillers together with standard error bars, corresponds to Figure 5.12. The Kenyan results for the subject relative clause fillers in Figure 5.19 at least generally follow the same trend as in the British English study: simple Ø-subject relative clauses receive by far the lowest scores. Besides this, an increased complexity causes lower scores for the grammatical fillers with *that* and *who*. Finally, simple *whom*-subjects get scores between the grammatical fillers and the simple Ø-subjects.

Figure 5.19 also diverges from the British English results (the increase of scores for Ø-subjects in Ø-complementizer bridge structures is less pronounced than in British English; *whom*-subjects receive much higher scores in such clauses; cf. Figure 5.12). Once more this illustrates the need for further experimental studies. However, for the present analysis, only the simple *whom*- and Ø-subject fillers were kept as reference points (following the British English study).

The first context which will be investigated in more detail is preposition placement in simple relative clauses. The general distribution of pied-piping and stranding in Figure 5.20 resembles the one observed in the British

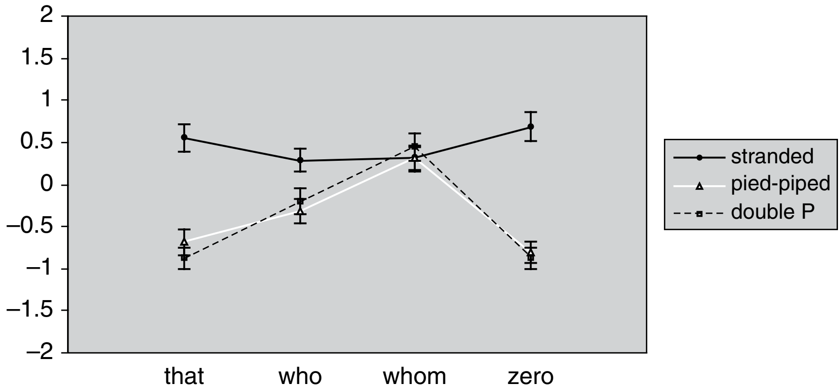

Figure 5.20 *Means for* PREPOSITION PLACEMENT *across* RELATIVIZERS *in simple relative clauses (Kenyan English data)*

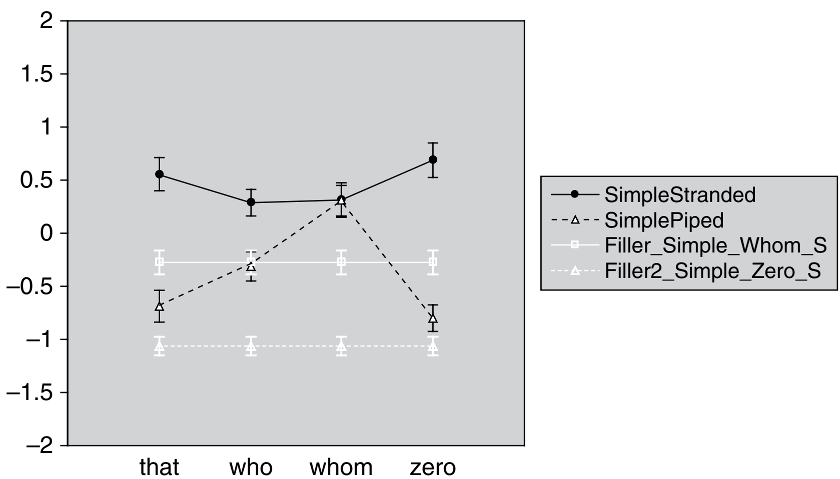

Figure 5.21 *Means for* PREPOSITION PLACEMENT *across* RELATIVIZERS *in simple relative clauses plus selected ungrammatical fillers (Kenyan English data)*

English study: stranding is preferred for all relativizers except *whom*, for which both preposition placement variants are considered equally good. Furthermore, compared to the British English results, *who*-relative clauses have indeed slightly lower scores than the ones containing Ø (cf. Figure 5.13). The most striking difference between the two studies, however, is the fact that – as can be seen in Figure 5.20 – even in simple *whom*-clauses doubled prepositions are judged equal to pied-piping.

Figure 5.21 then compares the judgements for pied-piping and stranding to the selected ungrammatical fillers. As Figure 5.21 shows, all simple

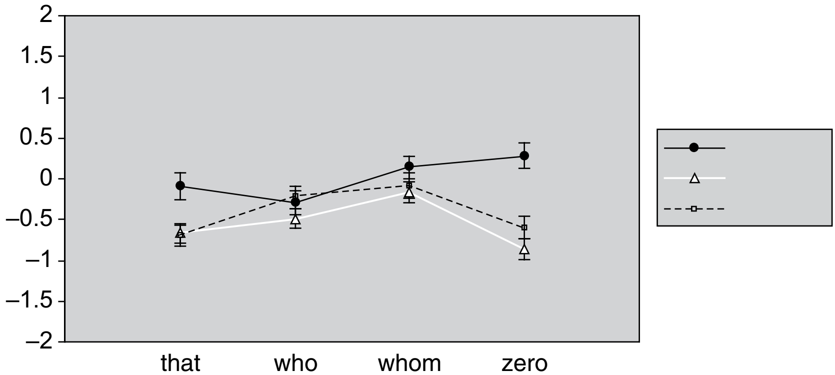

Figure 5.22 *Means for* PREPOSITION PLACEMENT *across* RELATIVIZERS *in complex* that-*complementizer relative clauses (Kenyan English data)*

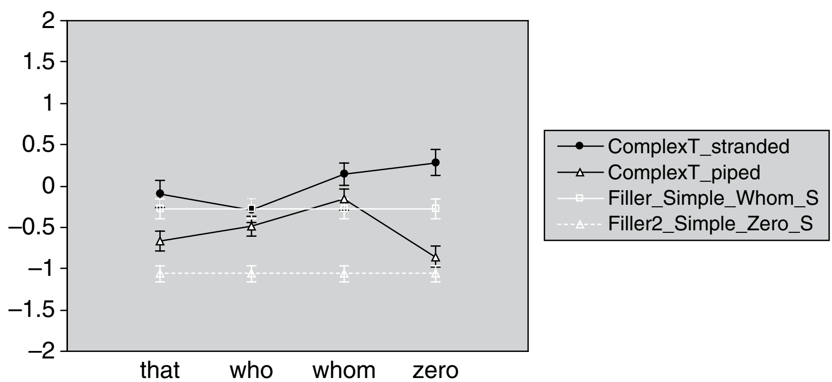

Figure 5.23 *Means for* PREPOSITION PLACEMENT *across* RELATIVIZERS *in complex* that-*complementizer relative clauses plus selected ungrammatical fillers (Kenyan English data)*

stranded relative clauses and *whom*-pied-piped stimuli are judged better than the ungrammatical *whom*-subject fillers. In addition to this, P + *who* structures receive scores similar to the *whom*-subject fillers. P + *that*- and P + Ø-relative clauses, on the other hand, get judgements worse than the case mismatch fillers but have higher scores than the Ø-subject fillers (see below for a statistical test of these differences).

Remember that in British English a *that*-complementizer bridge structure (cf. Figure 5.15) led to a decrease in judgement scores for stranding but this variant nevertheless was clearly preferred over the other two options. Figure 5.22 indicates that the situation in the Kenyan English data is somewhat more complex. The graph demonstrates that for *that* and Ø stranding is clearly preferred in the most complex clausal context, i.e. *that*- complementizer bridge verbs. Moreover, *who*, *whom* and Ø appear to yield better judgements for preposition-doubling than pied-piping.

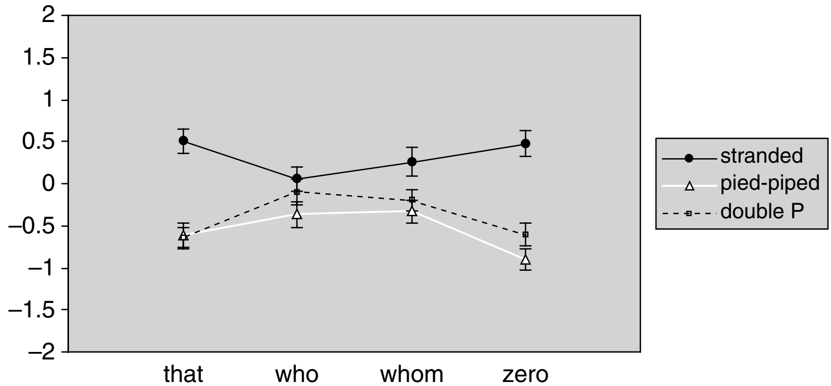

Figure 5.24 *Means for* PREPOSITION PLACEMENT *across* RELATIVIZERS *in complex Ø-complementizer relative clauses (Kenyan English data)*

Figure 5.23 then plots the judgement score for the stranded and the pied-piped complex *that*-clauses together with the ungrammatical filler set. As in the simple clauses, pied-piped *that*- and Ø- relative clauses have scores higher than the ungrammatical Ø-subject fillers. In addition to this, the difference between P + *that* and P + *who* seems less pronounced than in British English (cf. Figure 5.16). Finally, also note that *who* + P has the lowest scores out of all relativizers in stranded contexts, patterning with *whom*-subject fillers, while stranding with Ø-relativizers clearly yield results better than these ungrammatical fillers.

Finally, Figure 5.24 investigates the distribution of judgements in relative clauses containing Ø-complementizer bridge structures. Similar to bridge structures with *that*-complementizers, with Ø-complementizers *who*, *whom* and Ø give apparently better judgements for preposition-doubling than pied-piping. Likewise stranding also gives an analogous distribution with *who* + P structures receiving the lowest scores and Ø + P receiving the highest scores.

The inclusion of the ungrammatical fillers then makes it possible to investigate the effects for pied-piping and stranding more closely. Just as in the simple relative clauses, stranding once more yields judgements better than the *whom*-subject fillers in Figure 5.25. Pied-piping with *who* and *whom*, on the other hand, receives judgements comparable to these filler types.

In all of the above graphs which plot the results for pied-piping together with the ungrammatical filler set, P + *that* seems to have been rated better than the ungrammatical Ø-subject fillers, approximating the scores for P + *who* (a result supported by the Tukey test in Table 5.11). Yet, once the difference between the two P + overt relativizer structures and the Ø-subject fillers is investigated statistically a significantly different effect surfaces. In Table 5.12 it can be observed that P + *who* and P + *that* indeed have judgements higher than the Ø-subject fillers, but that this difference in means is only significant for the *who*-data.

Table 5.12 *T-test results P +* who/that-*relativizer vs simple Ø-subject fillers*

Complexity	P + X	*filler	Mean difference	*T*-test
simple	P + *who*	*simple Ø-subject	0.745	$t(35) = 4.180$, $p < 0.001$
	P + *that*	*simple Ø-subject	0.376	$t(35) = 2.019$, $p > 0.05$
that-complex	P + *who*	*simple Ø-subject	0.575	$t(35) = 3.253$, $p < 0.005$
	P + *that*	*simple Ø-subject	0.394	$t(35) = 2.382$, $p > 0.01$
Ø-complex	P + *who*	*simple Ø-subject	0.694	$t(35) = 3.927$, $p < 0.001$
	P + *that*	*simple Ø-subject	0.448	$t(35) = 2.314$, $p > 0.01$

Bonferroni-adjusted p-level: $p = 0.05/6 = 0.0083$. Rows with significant effects are shaded in grey.

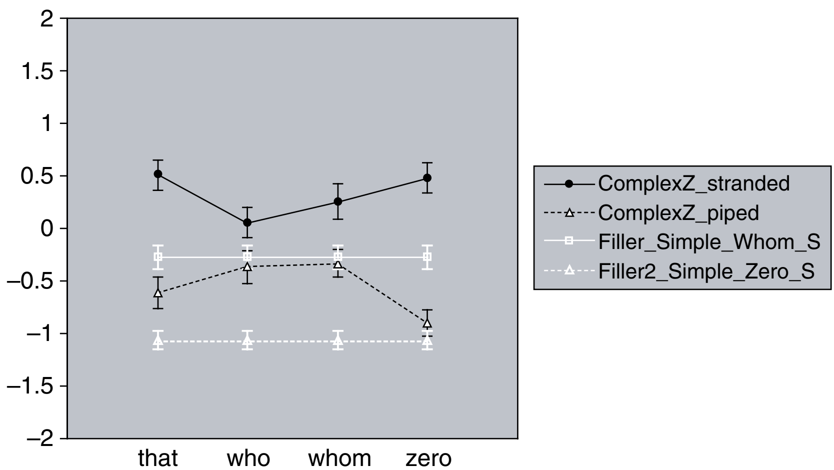

Figure 5.25 *Means for* PREPOSITION PLACEMENT *across* RELATIVIZERS *in complex Ø-complementizer relative clauses plus selected ungrammatical fillers (Kenyan English data)*

The statistical analysis of the Kenyan data thus also proves P + *that* (and consequently P + *Ø*, which always has lower scores) a hard grammatical constraint violation similar to simple *Ø*-subject constructions. Moreover, in line with the processing advantage associated with such structures, Kenyan speakers generally judge preposition-doubling at least as good as or better than pied-piping. Furthermore, regardless of the level of complexity, stranding is preferred over pied-piping with prepositional verbs (though a *that*-complementizer bridge structure significantly leads to lower scores), with a

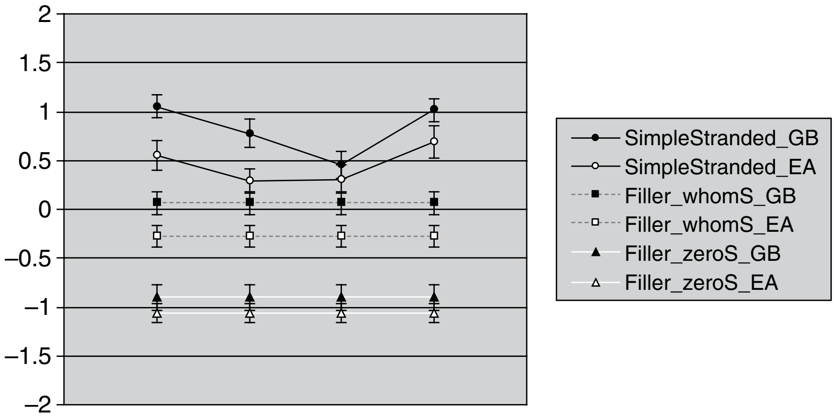

Figure 5.26 *Means for simple stranded relative clauses plus selected ungrammatical fillers (British and Kenyan English data)*

slight preference of Ø + P over *who* + P structures. Finally, however, it needs to be stressed that the Kenyan data turned out to be much noisier than the British: while a model containing the factors PREPOSITION PLACEMENT and PREPOSITION PLACEMENT*RELATIVIZER explained 67 percent of the variation of the British data (section 5.2.1), it only covers a mere 16 percent ($\eta^2 = 0.12$ and 0.04, respectively) of the Kenyan data.

Comparing the effect of complexity on preposition placement separately for British and Kenyan English obviously had the advantage of investigating the observed results relative to the judgements of the ungrammatical fillers in the two varieties. This allowed a more meaningful interpretation of effects, since in Magnitude Estimation experiments judgements are always relative. As it turned out, the two varieties seemed to display roughly similar effects in that, for example, stranding is clearly favoured with prepositional verbs regardless of the complexity of a clause. Comparing the results for the experimental stimuli that received the highest scores in both studies, i.e. simple stranded relative clauses, however, also reveals a striking difference between the two varieties (Figure 5.26; note that the scores for the two studies can be compared since all judgements were converted to *z*-scores). Figure 5.26 also includes the scores for the ungrammatical simple *whom*- and Ø-subject fillers. As can be seen, Kenyan speakers generally gave lower scores for all the stimuli in the figure. Since the relative scores for all other experimental conditions followed similar trends in both studies, this means that the Kenyan English speakers generally judged all preposition placement conditions lower than the British English speakers. This is therefore a clear case of different degrees of entrenchment: the grammars of British English speakers are more deeply entrenched, leading to better scores for stimuli. Due to a lower degree of entrenchment, Kenyan speakers, on the other hand, give lower judgement scores for all conditions. Note, however, that for the clearly ungrammatical Ø-subject structures the judgements of the two groups converge.

5.3 Preposition placement in interrogative clauses

After the studies on preposition placement in relative clauses, the final experiment investigated PREPOSITION PLACEMENT (three levels: stranded; pied-piped; doubled) in interrogative clauses. As in the first study (section 5.1) the two other factors tested included PP TYPE (three levels: prepositional verbs; temporal/locative sentence adjuncts; manner-degree/frequency-duration adjuncts) and displaced element, i.e.INTERROGATIVE PRONOUN (two levels: *who* vs *whom; which* vs *what*; see below). The design therefore crossed PREPOSITION PLACEMENT × PP TYPE × INTERROGATIVE PRONOUN yielding 3 × 3 × 2 = 18 cells. In a first step, for each of the different PP TYPE conditions six different lexicalizations were used. As (5.12)–(5.14) shows, the same lexicalizations as in experiment one were employed (cf. 5.1)–(5.3):

(5.12) **Prepositional verbs**:
a. **(On)**$_1$ who$_1$/whom$_2$ did Jane rely **(on)**$_2$?
b. **(With)**$_1$ who$_1$/whom$_2$ did Claudia sleep **(with)**$_2$?
c. **(At)**$_1$ who$_1$/whom$_2$ did Bill laugh **(at)**$_2$?
d. **(About)**$_1$ who$_1$/whom$_2$ did Steve talk about? **(about)**$_2$.
e. **(For)**$_1$ who$_1$/whom$_2$ did Mike apologise? **(for)**$_2$.
f. **(Of)**$_2$ who$_1$/whom$_2$ did she dream **(of)**$_2$?

(5.13) **Locational/temporal adjunct PPs**:
a. **(In)**$_1$ which$_1$/what$_2$ room did the murderer kill the victim **(in)**$_2$?
b. **(At)**$_1$ which$_1$/what$_2$ time did the party start **(at)**$_2$?
c. **(On)**$_1$ which$_1$/what$_2$ island did he find gold **(on)**$_2$?
d. **(On)**$_1$ which$_1$/what$_2$ day did James arrive **(on)**$_2$?
e. **(In)**$_1$ which$_1$/what$_2$ year did Elvis die **(in)**$_2$?
f. **(At)**$_1$ which$_1$/what$_2$ bar did they enjoy a few cocktails **(at)**$_2$?

(5.14) **Manner-degree/frequency-duration adjunct PPs**:
a. **(With)**$_1$ which$_1$/what$_2$ precision did Ben work **(with)**$_2$?
b. **(In)**$_1$ which$_1$/what$_2$ way did he achieve his goal **(in)**$_2$?
c. **(With)**$_1$ which$_1$/what$_2$ ease did he win **(with)**$_2$?
d. **(During)** which$_1$/what$_2$ service did they sit **(during)**$_2$?
e. **(With)**$_1$ which$_1$/what$_2$ frequency did earthquakes occur **(with)**$_2$?
f. **(On)**$_1$ which$_1$/what$_2$ occasions did Kelly faint **(on)**$_2$?

In (5.12)–(5.14) the possible lexicalizations for the different PP types are given together with the two levels of the factor INTERROGATIVE PRONOUN. As the sentences show, the various PP types interact with the range of acceptable interrogative pronouns: *who* and *whom* are only possible with prepositional verbs, while the two adjunct PP types only license *which* and *what*. As can be seen, for the experiment *who* and *which* were grouped together as one level (as marked by the subscripted '1'), while *whom* and *what* were classified as the second level of the factor INTERROGATIVE PRONOUN (cf. the

Table 5.13 *Filler set 1*

Factor combination	Filler stimulus
Subj_who_Vpast	Who sued Janet?
Subj_what_Vpast	What angered Nina?
Subj_which_Vpast	Which student hugged the teacher?
Obj_who_did_Vpres	Who did the teenager like?
Obj_whom_did_Vpres	Whom did Mark call?
Obj_which_did_Vpres	Which problem did the dentist encounter?
*Obj_who_no_did_Vpres	Who John find?
*Obj_whom_no_did_Vpres	Whom Steve hit?
*Obj_what_no_did_Vpres	What kind of steak Jeff eat?
*Obj_who_did_Vpast	Who did Mary bought?
*Obj_whom_did_Vpast	Whom did David saw?
*Subj_whom_Vpast	Whom remembered Sally?

subscripted '2'). While this decision was linguistically well-motivated, this factor interaction will have to be kept in mind for the interpretation of the results below.

As in the two previous studies, token sets with all six possible PREPOSITION PLACEMENT × INTERROGATIVE PRONOUN conditions were created for all the lexicalizations in (5.12)–(5.14). These were then counterbalanced in the usual way, yielding six material sets which exhibited all eighteen factor combinations but never used the same lexicalization twice within a material set. To these experimental stimuli, thirty-six filler sentences were added (resulting in a complete set of fifty-four stimuli with a 2:1 filler : experimental stimuli ratio). The fillers included eighteen grammatical and eighteen ungrammatical sentences and were created as follows.

First, six grammatical tokens were created by the combination of three interrogatives in which the subject position was questioned (using three different *wh*-items, namely, *who*, *what* and *which*) and three clauses in which the object position was questioned (employing *who*, *whom* and *which*). In addition to this, six ungrammatical tokens were included, three in which an expected operator *did* was missing (e.g. *Whom Steve hit?*), two in which past-tense marking was doubled on the operator *do* and the main verb (e.g. *Whom **did** David **saw**?*) and one with a *whom*-subject (e.g. *Whom remembered Sally?*). The resulting twelve fillers are summarized in Table 5.13. The same procedure was repeated twice until the fillers consisted of three sets like the one in Table 5.13 (see Appendix A.5.3.1 for material set 1).

The Magnitude Estimation experiment was carried out by a printed questionnaire created by the WebExp software. Results were transformed to z-scores and subjected to repeated measures ANOVAs as before.

5.3.1 British English

For the experiment thirty native speakers of British English (twenty-four female, six male; age range 20–69) were recruited. These comprised twenty students and lecturers from the University of Central Lancashire, five British citizens working as lecturers or studying at the University of Regensburg and five speakers from North Walsham, Norfolk, with whom contact had been established via a mutual acquaintance. The between-subject factors again turned out non-significant (AGE at $F(1,28) = 2.874$, $p > 0.10$; GENDER at $F(1,28) = 0.025$, $p > 0.85$), while the significant within-subject factors were the following (for the full analysis, see Appendix A.5.3.2):

- PREPOSITION PLACEMENT:
  $F_1(2,58) = 20.637$, $p < 0.001$, $\eta^2 = 0.19$
  $F_2(2,10) = 19.346$, $p < 0.001$, $\eta^2 = 0.41$
- PP-TYPE:
  $F_1(2,58) = 14.902$, $p < 0.001$, $\eta^2 = 0.09$
  $F_2(2,10) = 25.900$, $p < 0.001$, $\eta^2 = 0.18$

as well as the interaction of

- PREPOSITION PLACEMENT*PP TYPE:
  $F_1(4,116) = 7.413$, $p < 0.001$, $\eta^2 = 0.06$
  $F_2(4,20) = 4.605$, $p < 0.01$, $\eta^2 = 0.12$

Note that the model with the variables PREPOSITION PLACEMENT, PP TYPE and PREPOSITION PLACEMENT*PP TYPE thus explains 34 percent of the by-subject variation ($\eta^2 = 0.19 + 0.09 + 0.06$, respectively) and 71 percent of the by-item variation ($\eta^2 = 0.41 + 0.18 + 0.12$, respectively).

The results of post-hoc Tukey test of this interaction effect are presented in Table 5.14. Just as expected by the corpus study and the first experiment on relative clauses, Table 5.14 proves that in interrogatives preposition-stranding is also considered worst with manner/frequency adjunct PPs. The difference between prepositional verbs and locational/temporal adjunct PPs, on the other hand, does not come out as significant for the stranded data. In contrast to this, in the pied-piped data locational/temporal adjunct PPs receive significantly higher scores than the other two PP types. This probably indicates that interrogative sequences such as *in which place* or *at what time* are more deeply entrenched, i.e. far more lexicalized, than the corresponding pied-piped prepositional verb or manner/frequency adjunct PP structures. Finally, doubled prepositions in interrogative clauses receive statistically equal scores regardless of the PP type involved.

Next these interaction effects were investigated by plotting the mean judgement scores with standard error bars for the experimental conditions together with those for the fillers. Since the type of interrogative pronoun depends on the PP type in question (*who/whom* for prepositional

Table 5.14 *Result of post-hoc Tukey test for* PREPOSITION PLACEMENT**PP* TYPE *interaction (British English data)*

Dependent variable	(I) PP type	(II) PP type	Mean difference	Significance
$P_{stranded}$	$V_{prepositional}$	$Adjunct_{loc/temp}$	−0.084	ns
		$Adjunct_{man/frequ}$	0.717	$p < 0.001$
	$Adjunct_{loc/temp}$	$V_{prepositional}$	0.084	ns
		$Adjunct_{man/frequ}$	0.801	$p < 0.001$
	$Adjunct_{man/frequ}$	$V_{prepositional}$	−0.717	$p < 0.001$
		$Adjunct_{loc/temp}$	−0.801	$p < 0.001$
$P_{pied\text{-}piped}$	$V_{prepositional}$	$Adjunct_{loc/temp}$	−0.551	$p < 0.001$
		$Adjunct_{man/frequ}$	−0.177	ns
	$Adjunct_{loc/temp}$	$V_{prepositional}$	0.551	$p < 0.001$
		$Adjunct_{man/frequ}$	0.374	$p < 0.05$
	$Adjunct_{man/frequ}$	$V_{prepositional}$	0.177	ns
		$Adjunct_{loc/temp}$	−0.374	$p < 0.05$
$P_{doubled}$	$V_{prepositional}$	$Adjunct_{loc/temp}$	−0.079	ns
		$Adjunct_{man/frequ}$	0.195	ns
	$Adjunct_{loc/temp}$	$V_{prepositional}$	0.079	ns
		$Adjunct_{man/frequ}$	0.274	ns
	$Adjunct_{man/frequ}$	$V_{prepositional}$	−0.195	ns
		$Adjunct_{loc/temp}$	−0.274	ns

Rows with significant effects are shaded in grey.

verbs, *which/what* for adjunct PPs; cf. above), the results for the two levels of INTERROGATIVE PRONOUN are given in individual graphs. Figure 5.27 therefore gives the results for *who* and *which* across the three PP TYPE conditions (and the mean scores for the ungrammatical fillers). As Figure 5.27 shows, doubled prepositions generally lead to scores lower than pied-piping or stranding. At the same time, doubled prepositions get scores higher than the ungrammatical fillers. This can be explained once more by the fact that locally both instances of the preposition in such constructions are acceptable. The violation caused by doubled preposition is therefore less severe than that of the ungrammatical fillers, which include local violations such as case mismatch (*whom*-subjects), tense-marking (tense inflection on the main verb in interrogatives that already have a tense-marked operator) and structural violations (i.e. a missing obligatory *do*-operator).

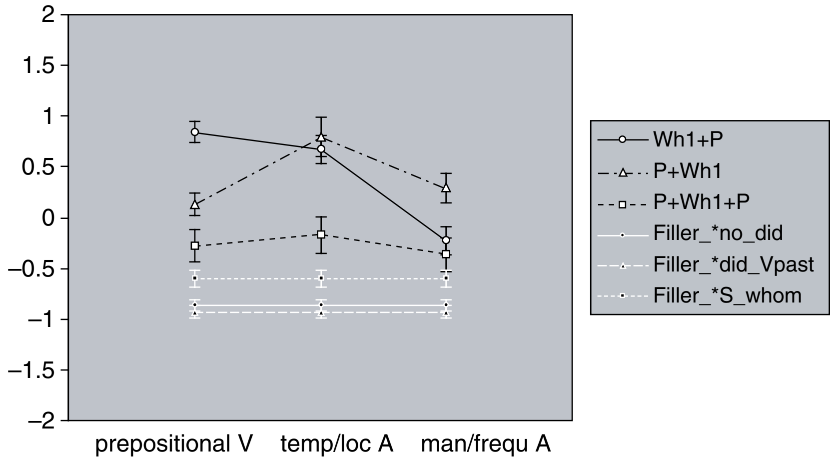

Figure 5.27 *Means for* PREPOSITION PLACEMENT**PP* TYPE *for interrogative pronoun* wh-*1 plus ungrammatical fillers (British English data)*

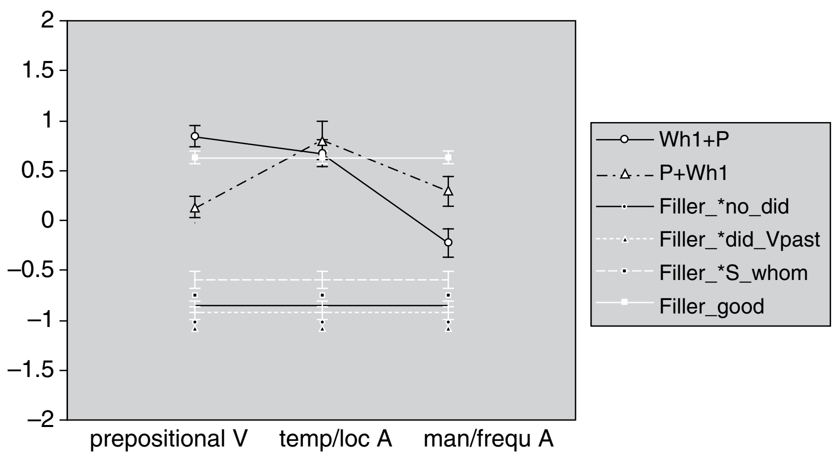

Figure 5.28 *Means for preposition stranding/pied-piping*PP* TYPE *for interrogative pronoun* wh-*1 plus all fillers (British English data)*

Turning to the two canonical preposition placement variants, Figure 5.27 gives pied-piping with manner/frequency adjuncts (P + *which*) as being preferred, while, as expected, stranding with these PPs significantly decreases acceptability scores. Note furthermore that it is only with manner/frequency adjuncts that the scores for stranding approximate those of preposition-doubling. With prepositional verbs, on the other hand, *who* + P is strongly preferred over P + *who*, but the latter nevertheless gets scores much higher than all ungrammatical fillers. Finally, for locational/temporal

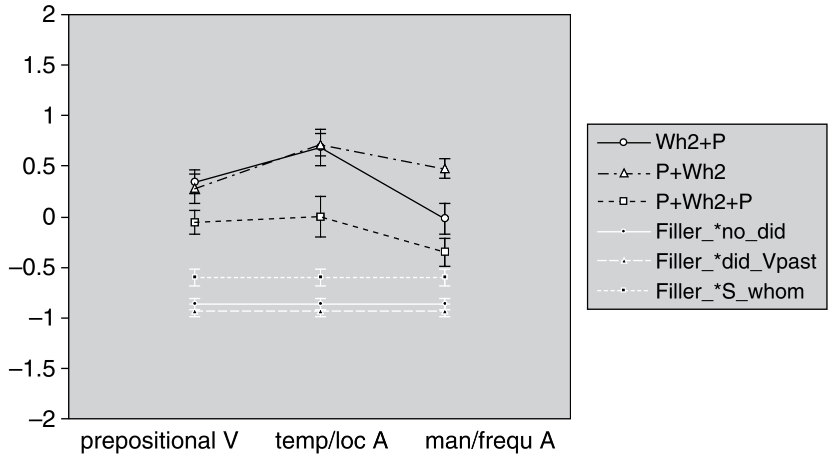

Figure 5.29 *Means for* PREPOSITION PLACEMENT*PP TYPE *for interrogative pronoun* wh-2 *plus ungrammatical fillers (British English data)*

adjunct PPs stranding (*which* + P) and pied-piping (P + *which*) give equally high acceptability scores.

Figure 5.28 then makes it possible to check the results for preposition-stranding and pied-piping for the first level of INTERROGATIVE PRONOUN against all fillers. In the figure, preposition stranding with prepositional verbs and locational/temporal adjunct PPs is better than or equal to the judgements for the grammatical fillers. Stranding with manner/frequency PPs, on the other hand, exhibits the distribution of a soft grammatical constraint: the judgement scores for this condition are well below the set of grammatical fillers but also better than the set of hard grammatical constraints exhibited by the ungrammatical fillers.

Next, the effects of the second group of interrogative pronouns (*whom* / *what*) are looked into. Following the distribution of the first set of interrogative pronouns, both *who* and *what* have scores for doubled prepositions in Figure 5.29 below the grammatical preposition placement variants, but higher than the ungrammatical fillers. Besides this, prepositional verbs with *whom* differ from those with *who* in Figure 5.27: while pied-piping with *whom* gets slightly better judgements, stranding with *whom* leads to a decrease in judgement scores. This indicates that *whom* might be less entrenched in interrogative clauses than in relative clauses, where no such effect was observed for British English (cf. section 5.2.1). The remaining effects are similar to those observed above: for locational/temporal adjunct PPs both stranding and pied-piping are equally good, while stranding with manner/frequency PPs yields much lower acceptability scores.

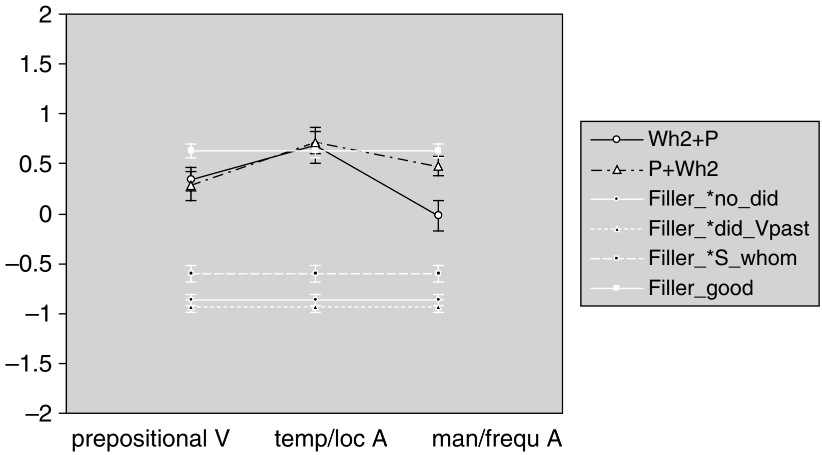

Figure 5.30 *Means for preposition stranding/pied-piping*PP* TYPE *for interrogative pronoun* wh-*2 plus all fillers (British English data)*

Figure 5.30 once more compares the judgements for pied-piping and stranding with the second set of pronouns with all grammatical and ungrammatical fillers. Both P + *whom* and *whom* + P have judgements below those for the grammatical fillers, supporting the claim that *whom* generally might be less entrenched in interrogative clauses. As expected, the only context in which stranding in Figure 5.30 is judged unacceptable is again the manner/frequency adjunct PPs: the judgement scores of these lie between the set of grammatical fillers and the ungrammatical fillers violating hard grammatical constraints.

For British English, this experiment thus corroborated again the status of preposition-stranding with manner/frequency adjunct PPs as a soft grammatical constraint violation, regardless of the clause type involved. In addition to this, preposition-stranding and pied-piping of locational/temporal adjunct PPs with both *which* and *what* were judged equally good within the range of the grammatical fillers. This contrasts with the situation in relative clauses, where the experiment in section 5.1.1 had identified these PPs as preferring pied-piping. Just as in the corpus results, this confirms the hypothesis that interrogatives are less complex clausal environments than relatives and consequently license stranding more easily – provided the stranded preposition can be thematically interpreted as an entity by the semantic frame of the main predicate. The lower degree of processing effort associated with interrogative clauses also explains why doubled prepositions have clearly lower scores than pied-piping in this experiment: in the relative clause study (cf. section 5.2.1) it was shown that in the most complex clausal context doubled prepositions facilitate processing, which leads to them receiving scores comparable to pied-piped prepositions. In the simpler interrogative clauses there is no need to reduce processing cost in a similar way, leading to the doubled-preposition structure standing out as a conspicuous non-standard construction (though the local acceptability of the two instances of the preposition still ensures scores higher than the set of ungrammatical fillers).

Finally, what exactly is the status of the P + *who* structures? How do these structures compare statistically to the set of grammatical and ungrammatical fillers? As it turns out, P + *who* does in fact receive judgements which are significantly below those of the grammatical fillers ($t(29) = 4.203$, $p < 0.001$). On the other hand, it also has scores significantly higher ($t(29) = 5.364$, $p < 0.001$) than the ungrammatical *whom*-subject fillers, which had been identified as the ungrammatical fillers with the highest mean scores. Thus, while in simple relative clauses P + *who* was judged worse than ungrammatical *whom*-subjects (cf. Figure 5.14), in interrogatives the structure is considered better than the case mismatch violation. The reason for this is obviously that, unlike relative clauses, interrogatives have *in situ* uses of *who* as the complement of prepositions (e.g. *He did that to who*?). Consequently, the P + *who* structure is already to a certain degree entrenched in interrogative clauses, which explains its increased judgement scores in these contexts.

5.3.2 Kenyan English

During my field trip to Kenya in September 2006, the interrogative experiment was also carried out with thirty students of the University of Nairobi (thirteen female, sixteen male;[6] age range 21–22). The low number of speakers of a non-Bantu language (three Dholuo, one Kalenjin speakers) once more prevented investigating the factor first-language.

AGE and GENDER were again non-significant at $F(1,24) = 0.305$, $p > 0.55$ and $F(1,24) = 1.135$, $p > 0.25$, respectively. In contrast to this, significant effects turn out to be the following within-subject factors (the full analysis is provided in Appendix A.5.3.3):

- PREPOSITION PLACEMENT:
  $F_1(2,58) = 33.290$, $p < 0.001$, $\eta^2 = 0.17$
  $F_2(2,10) = 46.053$, $p < 0.001$, $\eta^2 = 0.43$
- PP TYPE (only significant by-item):
  $F_1(2,58) = 2.964$, $p > 0.05$, $\eta^2 = 0.02$
  $F_2(2,10) = 4.267$, $p < 0.05$, $\eta^2 = 0.05$

as well as the interactions of

- PREPOSITION PLACEMENT*PP TYPE:
  $F_1(4,116) = 8.275$, $p < 0.001$, $\eta^2 = 0.08$
  $F_2(4,20) = 5.665$, $p < 0.005$, $\eta^2 = 0.19$

[6] One speaker failed to indicate his or her gender.

Table 5.15 *Result of post-hoc Tukey test for* PREPOSITION PLACEMENT**PP* TYPE *interaction (Kenyan English data)*

Dependent variable	(I) PP type	(II) PP type	Mean difference	Significance
$P_{stranded}$	$V_{prepositional}$	$Adjunct_{loc/temp}$	0.312	ns
		$Adjunct_{man/frequ}$	0.508	$p < 0.005$
	$Adjunct_{loc/temp}$	$V_{prepositional}$	−0.312	ns
		$Adjunct_{man/frequ}$	0.196	ns
	$Adjunct_{man/frequ}$	$V_{prepositional}$	−0.508	$p < 0.005$
		$Adjunct_{loc/temp}$	−0.196	ns
$P_{pied\text{-}piped}$	$V_{prepositional}$	$Adjunct_{loc/temp}$	−0.662	$p < 0.001$
		$Adjunct_{man/frequ}$	−0.152	ns
	$Adjunct_{loc/temp}$	$V_{prepositional}$	0.662	$p < 0.001$
		$Adjunct_{man/frequ}$	0.510	$p < 0.005$
	$Adjunct_{man/frequ}$	$V_{prepositional}$	0.152	ns
		$Adjunct_{loc/temp}$	−0.510	$p < 0.005$
$P_{doubled}$	$V_{prepositional}$	$Adjunct_{loc/temp}$	0.318	ns
		$Adjunct_{man/frequ}$	0.247	ns
	$Adjunct_{loc/temp}$	$V_{prepositional}$	−0.318	ns
		$Adjunct_{man/frequ}$	−0.071	ns
	$Adjunct_{man/frequ}$	$V_{prepositional}$	−0.247	ns
		$Adjunct_{loc/temp}$	0.071	ns

Rows with significant effects are shaded in grey.

- PP TYPE*INTERROGATIVE PRONOUN:
  $F_1(1.545,44.792) = 3.572$, $p < 0.05$, $\eta^2 = 0.01$
  $F_2(2,10) = 5,416$, $p < 0.05$, $\eta^2 = 0.04$

Note that a model corresponding to the one for the British data (cf. section 5.3.1) with the variables PREPOSITION PLACEMENT, PP TYPE and PREPOSITION PLACEMENT*PP TYPE thus explains 27 percent of the by-subject variation ($\eta^2 = 0.17 + 0.02 + 0.08$, respectively) and 67 percent of the by-item variation ($\eta^2 = 0.43 + 0.05 + 0.19$, respectively).

Next, the cause of the interaction effect involving the factor PREPOSITION PLACEMENT was examined in a post-hoc Tukey test (Table 5.15). Just as in the British English study (cf. Table 5.14), the factor PP TYPE has no effect on the judgements of doubled prepositions. Furthermore, in the pied-piped data locational/temporal adjunct PPs are again identified as significantly better than the other two PP types. Finally, for the stranded manner/frequency

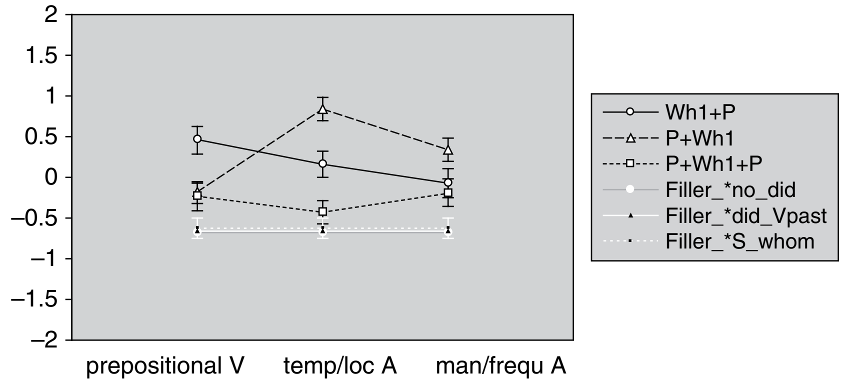

Figure 5.31 *Means for* PREPOSITION PLACEMENT**PP* TYPE *for interrogative pronoun* wh-*1 plus ungrammatical fillers (Kenyan English data)*

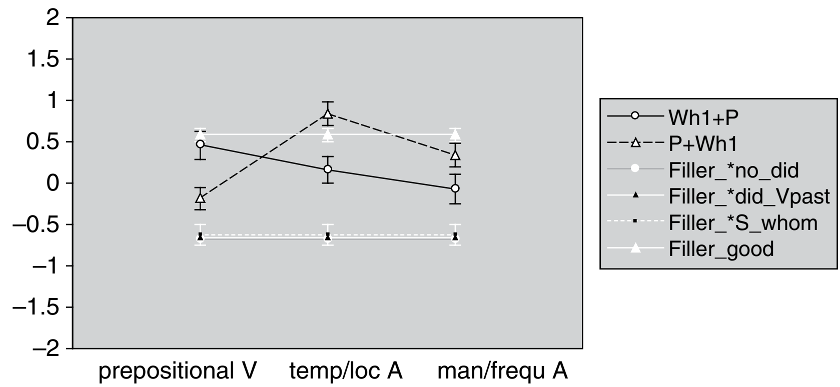

Figure 5.32 *Means for preposition stranding/pied-piping*PP* TYPE *for interrogative pronoun* wh-*1 plus all fillers (Kenyan English data)*

adjunct PPs the expected soft constraint violation appears to operate, leading to significantly worse scores than for stranding with prepositional verbs (though apparently not with locational/temporal adjunct PPs, but cf. below).

As before, the two levels of the factor INTERROGATIVE PRONOUN will be investigated independently. Figure 5.31 therefore gives the results for *who* and *which* across the three PP TYPES conditions (and the mean scores for the ungrammatical fillers). The first notable difference to the British English results concerns the status of the ungrammatical fillers: British speakers had differentiated between *whom*-subject, on the one hand, and missing-*did* and double tense-marked fillers, on the other hand, consistently assigning better judgements to the former (cf. Figure 5.27). The Kenyan speakers, however,

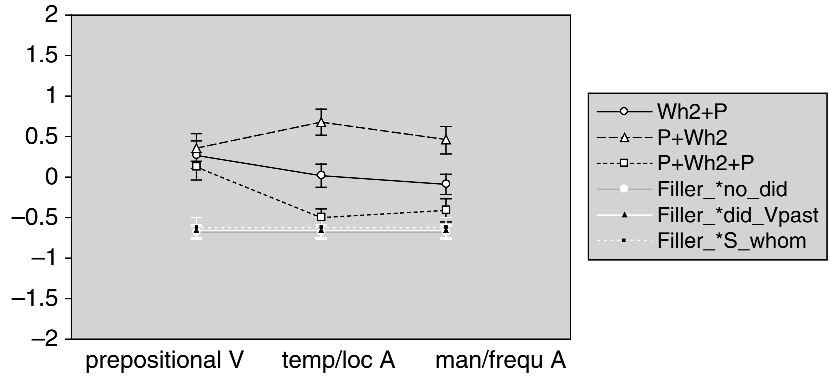

Figure 5.33 *Means for* PREPOSITION PLACEMENT**PP* TYPE *for interrogative pronoun* wh-*2 plus ungrammatical fillers (Kenyan English data)*

judged all three types of ungrammatical fillers as equally bad. In addition to that, it is not only for the stranded manner/frequency adjunct PPs that the judgements for preposition-stranding and doubling approximate each other. In Figure 5.31, for prepositional verbs P + *who* and P + *who* + P also have similar scores.

The effects of stranded and pied-piped prepositions across PP types is best illustrated by including the grammatical fillers. As Figure 5.32 shows, preposition-stranding in the Kenyan data resembles more closely the cline exhibited by the relative clause data in sections 5.1.1 and 5.1.2 (cf. Figures 5.4 and 5.8): stranding gets the best scores with prepositional verbs, locational/temporal adjunct PPs yield lower judgements and manner/frequency adjunct PPs have the lowest acceptability scores. While in British English locational/temporal adjunct PPs thus license both stranding and pied-piping in interrogatives, the Kenyan English data once more indicate the strong preference for pied-piping with adjunct PPs in this variety. Furthermore, while pied-piping is judged similarly to doubling with *who* (in the case of prepositional verbs), it is still deemed better than the ungrammatical filler set.

Next, Figure 5.33 provides an overview of preposition placement with *whom*- and *what*-interrogative pronouns. By far the most interesting effect in Figure 5.33 concerns doubled prepositions with prepositional verbs: P + *whom* + P structures are judged similar to stranded and pied-piped *whom*-structures, while in all other contexts doubled prepositions have significantly lower scores than the canonical preposition placement variants. Remember, however, that Kenyan speakers always prefer pied-piping with adjunct PPs, so the second, stranded instance of the preposition causes doubled prepositions with locational/temporal and manner/frequency adjuncts to receive low scores in Figures 5.31 and 5.33. In addition to this, in Figure 5.31 prepositional verbs co-occur with a *who*-pronoun, for which pied-piping alone

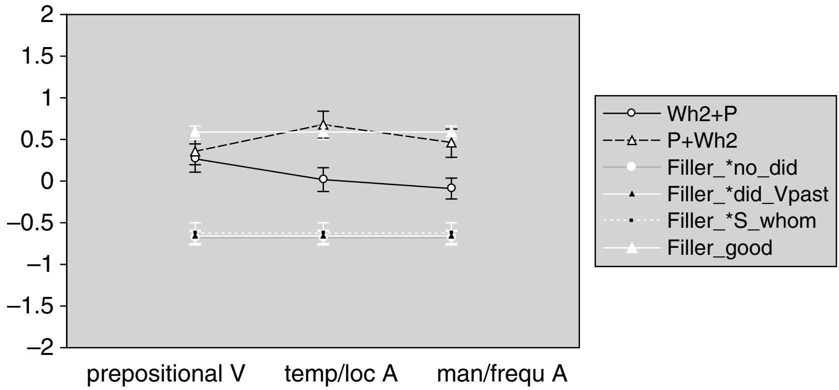

Figure 5.34 *Means for preposition stranding/pied-piping*PP* TYPE *for interrogative pronoun* wh-*2 plus all fillers (Kenya English data)*

already leads to low acceptability. It is thus only with *whom* that stranding and pied-piping are independently viable options, and in such a case the second-language learners give higher ratings for a structure that facilitates processing. This constitutes a notable divergence from the results of the first-language speakers. Since interrogative clauses are comparatively easy to process and preposition placement is so deeply entrenched in their mental grammars, British English speakers assign low values to doubled prepositions. Since the processing complexity of interrogatives per se is similarly low for Kenyan English speakers, the increased acceptability of doubled prepositions again highlights the lower degree of entrenchment of preposition placement in their grammars.

Finally, Figure 5.34 focuses on preposition-stranding and pied-piping with *whom* and *what*, contrasting the two options with all grammatical and ungrammatical fillers. As before, pied-piping is preferred for both adjunct PP types, which receive scores comparable to the grammatical fillers. Moreover, stranding decreases from prepositional verbs (which are still in the range of the grammatical fillers) to manner/frequency adjunct PPs, with locational/temporal adjunct PPs having intermediate scores.

This study has once more illustrated that Kenyan English speakers generally prefer pied-piping with adjunct PPs. Furthermore, the soft constraint violation of stranding with manner/frequency adjunct PPs has also been corroborated again. In addition to this, the second-language learners exhibited a preference for doubling with *whom* that was not observed with British speakers. This was deemed a reflection of the overall lower degree of entrenchment of preposition placement in the mental grammars of the Kenyan speakers. Finally, as in the British study, the question arises as to how to interpret the judgements of P + *who*, which end up somewhat in between the grammatical and ungrammatical fillers. Testing the P + *who* acceptability scores against the grammatical fillers clearly proves that the latter are significantly judged better ($t(29) = 5.029$, $p < 0.001$). Upon comparing P + *who* with the ungrammatical *whom*-subject fillers, a *t*-test shows that the higher scores of pied-piping with *who* are slightly significant ($t(29) = 2.238$, $p < 0.05$). Yet two *t*-tests were carried out on the same data set, which means that the *p*-values must be adjusted accordingly. The Bonferroni-adjusted *p*-value for these tests is 0.025 (=0.05/2), which means that only the first one can be said to measure a significant difference in means. The result of the latter (exact $p = 0.0331$) then does not qualify as significant, meaning that P + *who* structures cannot be said to be statistically better than the ungrammatical fillers. Tentatively this result might be caused by a lower degree of entrenchment of *in situ* P + *who* sequences in Kenyan English interrogatives.

5.4 Summary

In this chapter the results of three different sets of experiments on preposition placement were presented. As it turned out the following important effects emerged:

(1) Pied-piping with *that*- and Ø-relativizers is a hard grammatical constraint in British and Kenyan English (section 5.1 and 5.2).
(2) Pied-piping with *who*-relativizers does give lower acceptability scores than with *whom*, but is still considered significantly better than pied-piping with *that*- and Ø-relativizers in both varieties (section 5.2). While P + *who* is a case match violation, P + *that* is a qualitatively different phenomenon (i.e. a structure not provided for by the grammar).
(3) In the corpus study, the distribution of manner/frequency adjunct PPs indicated that preposition-stranding is not possible with such PPs, which do not add simple thematic participants to predicates. This claim has been corroborated for both varieties experimentally (in sections 5.1 and 5.3) and, in addition, has been identified as a soft constraint violation.
(4) The two varieties exhibited differences with respect to the stranding of locational/temporal adjunct PPs: while these received acceptable ratings in relative (5.1.1) and interrogative (5.3.1) clauses by the British English speakers, the Kenyan speakers strongly preferred pied-piping in these contexts. Preposition-stranding in Kenyan English is thus less productive and limited to fewer contexts than in British English. This was taken as a sign of a lower degree of entrenchment of preposition-stranding in Kenyan English in general (i.e. its strong association with lexically stored verb–preposition structures).
(5) With increasing complexity (of relative clauses), the stranding preference for prepositional verbs is increased in both varieties (sections 5.2.1.

and 5.2.2) since this ensures the correct interpretation of the lexicalized verb–preposition structures. The Kenyan data also exhibited two further complexity effects: first of all, the judgement scores for all experimental stimuli turned out to be lower than the corresponding ones made by the British English informants, highlighting again the lower degree of entrenchment of preposition placement in Kenyan English. Secondly, this lower degree of entrenchment together with the increasing clausal complexity led to increased judgement scores of doubled prepositions by the second-language speakers. This was attributed to the fact that doubled prepositions with prepositional verbs greatly facilitate processing (section 5.2.1).

In the next chapter, the above experimental results together with the ones obtained from the ICE corpora studies (chapter 4) will become the basis for a grammatical description of preposition placement in British and Kenyan English. As I will try to show, the syntactic theory known as Construction Grammar provides a framework that meets the criterion of explanatory adequacy while at the same time making it possible to account for the great many empirical findings unearthed in chapters 4 and 5.

6 Preposition placement: The case for a Construction Grammar account

6.1 Construction Grammar: A descriptively and explanatorily adequate linguistic theory

The previous two chapters were an attempt at providing an observationally adequate description of preposition placement in British and Kenyan English: employing empirical methods, various language-internal and -external factors were statistically identified as affecting preposition- stranding and pied-piping. Most modern theories of linguistics, however, do not stop at merely charting the distribution of grammatical phenomena. Instead, it is commonly accepted that all linguistic performance is the product of a cognitive mental system. Accordingly, the two main goals of such theories are to describe this mental system (also known as a speaker's competence) and to explain how children can acquire it. If a grammar is successful at providing a description of the linguistic competence of speakers it is labelled 'descriptively adequate' (Chomsky 1965: 24). If it furthermore gives a correct account of the language acquisition process, a grammar can be also said to be 'explanatorily adequate' (Chomsky 1965: 25–6).

One well-known framework aiming at these goals is Noam Chomsky's Principles-and-Parameters theory (for the various stages from Standard Theory to Minimalist Program, see Chomsky 1965, 1981, 1995, 2001). Assuming the existence of an innate language faculty (called Universal Grammar), proponents of this theory claim to have achieved a high level of descriptive and explanatory adequacy. Yet, as, for example, Culicover and Jackendoff (cf. 2005: 25–37) have pointed out, this alleged success is only achieved by ignoring a great number of linguistic phenomena. The standard approach of this research paradigm is 'to focus on the core system, putting aside phenomena that result from historical accident, dialect mixture, personal idiosyncrasies, and the like' (Chomsky 1995: 20). As it turns out, however, this results in a great many phenomena – all of which must be part of a speaker's mental representation of language – which cannot be accounted for by the Principles-and-Parameters theory.

One such phenomenon, for examples, are idioms like *kick the bucket* 'to die' or *spill the beans* 'to divulge information', i.e. lexical items which are larger than words and which, due to their non-conventional meaning, must be stored in the mental lexicon (see Croft and Cruse 2004: 248–53; Jackendoff 2002: 167–8). Both of these idioms are usually excluded from linguistic analysis in Principles-and-Parameters approaches since they are considered language-specific idiosyncrasies. Yet even these idiomatic lexical items participate to a certain degree in syntactic processes: both have an open subject-slot that can be filled by any (animate) NP (cf. *He / Bill / Our next door neighbour kicked the bucket*) and a main verb that inflects for tense and agreement (cf. *She spilt / will spill / has spilt the beans*). In addition to that, *spill the beans* can even be passivized (*The beans have been spilt*). In fact, there are some idioms that only appear in passives (e.g. NP *has it made*, NP *is fit to be tied*; see Jackendoff 2002: 171). These idioms are therefore clearly affected by normal syntactic processes and not just frozen expressions created by historical accident. Consequently these must be incorporated into any linguistic description, and cannot just be ignored as peripheral phenomena.

In addition to these idioms, English contains many more such 'peripheral' items: there are idiomatic structures that not only have an open subject-slot, but also contain VP-internal slots (e.g. *take* NP *to task*, *take* NP *for granted*, *show* NP *the door*; see Jackendoff 2002: 168). Besides this, there are also idioms that do not contain any open slots, but are entirely filled with lexical material (e.g. *the jig is up*, *that's the way the cookie crumbles*; see Jackendoff 2002: 169). All of these examples have a non-conventional meaning, again suggesting that these expressions must be stored in the lexicon (and are relegated to the periphery bin in Principles-and-Parameters approaches).

Finally, English even has some fairly frequent expressions that seem to violate principles of English syntax. An example is the 'covariational-conditional construction' (Goldberg 2003: 220; Jackendoff 2002: 178):

(6.1) a. the louder, the better
b. the more you hear, the less you care

From a synchronic perspective the structures in (6.1) are pretty odd since they consist of a determiner followed by a comparative phrase (6.1a), which can even be augmented by a clause (6.1b). Normally in English, however, determiners only combine with nouns. Moreover, the meaning of the second part (*the better*) depends on the first (*the louder*).

All of the above structures pose serious problems for Principles-and-Parameters approaches (for several others, cf. Croft and Cruse 2004: 225–56; Goldberg 2006: 5–18; Jackendoff 2002: 167–82). Yet, English has numerous such peripheral phenomena that 'are inextricably interwoven with the "core"' (Culicover and Jackendoff 2005: 26) and which somehow must be acquired by children. As such, ignoring the vast number of idiomatic expressions leads to a grammatical analysis that is neither observationally nor descriptively and explanatorily adequate.

Table 6.1 *The syntax–lexicon continuum (adapted from Croft and Cruse 2004: 255; incorporating arguments from Goldberg 2003: 220 and Jackendoff 2002: 176)*

Construction type		Traditional name	Example
Complexity	Schematicity		
complex	schematic	syntax	[Sbj Verb-tns Obj$_1$ Obj$_2$]
complex	(mostly) schematic	syntax	[Sbj *be*-tns Verb-*en* (PP$_{by}$)]
			[*the* Xer *the* Yer]
complex	substantive verb	subcat. frame	[Sbj *consume* Obj]
complex	(mostly) substantive	idiom	[kick-tns the bucket]
			[spill-tns the beans]
complex	substantive	idiom	[That's the way the cookie crumbles]
complex	schematic bound	morphology	[Noun-plural], [Verb-tns]
atomic	schematic	syntacic category	[Noun], [Adj]
atomic	substantive	word/lexicon	[green], [anaconda]

Instead of selectively focusing on 'core' aspects only, in the past few years an alternative approach to the Principles-and-Parameters theory has gained importance, which aims 'to account for the full range of facts about language' (Goldberg 2003: 219). These so-called 'Construction Grammar' approaches (which include Croft 2001; Fillmore and Kay 1996; Ginzburg and Sag 2000; Goldberg 2003; Jackendoff 2002; Langacker 2005; Tomasello 2003) claim that that there is 'a **uniform representation of all grammatical knowledge** in the speaker's mind, in the form of … constructions' (Croft and Cruise 2004: 255; original emphasis). Constructions in this sense of the word refer to all 'LEARNED PAIRINGS OF FORM WITH SEMANTIC OR DISCOURSE FUNCTION, including morphemes or words, idioms, partially lexically filled and fully general phrasal patterns' (Goldberg 2006: 5). Table 6.1 provides a few examples that illustrate how all levels of description involve constructions of varying size and complexity (note that different researchers might employ different formalizations of the phenomena in the table; more on this below).

There are two main parameters along which the constructions in Table 6.1 have been classified: with respect to their complexity, constructions are said to be atomic if they 'cannot be further divided into meaningful parts' (Croft and Cruse 2004: 255), otherwise they are considered complex. The degree of schematicity then depends on whether a position of a construction is specified with respect to its phonology: a fixed phonological form means the presence of a substantive element; an open slot that can be filled by various items is labelled schematic.

As Table 6.1 shows, all levels of grammatical description (given by their traditional names) can be captured within a Construction Grammar account, yielding a lexicon–syntax continuum: at the lexicon end, there are atomic, substantive items such as *green* or *anaconda*, which ever since Saussure have been classified as form–meaning pairings. Syntactic categories such as Noun or Adjective, on the other hand, are also atomic constructions but are schematic, i.e. they are abstract slots that can be filled by a whole range of phonological forms. Accordingly, these also have a more abstract meaning side: Nouns prototypically denote ‘things’, while the meaning of Adjectives normally involves ‘stative relations’ (see Broccias 2006: 86–90; Croft 2001: 84–107 and Langacker 2005: 113–28 for a much more detailed discussion).

At the other end of the scale, one finds complex, completely schematic entities such as the ditransitive construction [Sbj Verb-tns Obj_1 Obj_2], which as a whole has a meaning of ‘transfer or “giving”’ (Goldberg 2006: 9), and whose individual schematic slots are associated with particular semantic functions (the Subject slot is identified as the agent, and Object_1 and Object_2 encode the roles recipient and theme, respectively). This abstract construction then licenses prototypical sentences such as *Brad gave Angelina a ring* but also instantiates less typical examples like *Brad baked Angelina a cake* (see Goldberg 2003: 220–1, 2006: 9–10).

In between these two poles lie all of the above ‘peripheral’ phenomena: idioms such as *kick the bucket* and *spill the beans* are both complex constructions with most of their syntactic slots being associated with a fixed phonological form. Yet, while the former only as a whole has the meaning ‘to die’, in the latter the two individual substantive elements *spill* and *the beans* are associated with the non-conventional meaning of ‘divulge’ and ‘information’. The idiomatic ‘covariational-conditional construction’ [*the* Xer *the* Yer], on the other hand, is a largely schematic complex construction, whose meaning is ‘interpreted as involving an independent variable (identified by the first phrase [X]) and a dependent variable (identified by the second phrase [Y])’ (Goldberg 2003: 220). Finally, the sentence *That’s the way the cookie crumbles* is a complex, fully substantive construction whose meaning can roughly be paraphrased as ‘that’s how it is and you can’t do anything about it’.

Construction Grammar approaches can thus capture all kinds of idiosyncratic, ‘peripheral’ phenomena of a language. In addition to this, however, it is also possible to capture all of the perfectly compositional ‘core’ phenomena such as the ditransitive or the passive (Table 6.1) construction (though more on the latter below). In addition to this, the information in Table 6.1 is not taken to be stored mentally as a list of unrelated constructions. Instead, all versions of Construction Grammar agree that the constructions of a language form a structured inventory, which can be represented by taxonomic networks (cf. Croft and Cruse 2004: 262–5). Figure 6.1 shows how this structured inventory combines to create a sentence like *John spills the beans* (again different Construction Grammar approaches use different formalisms; Figure 6.1 is therefore intended as an expository example only). As argued above, the construction [*spill* [*the beans*]] must be stored since its two elements have a non-conventional meaning which they only exhibit in this particular idiom. In fact, most Construction Grammar theories

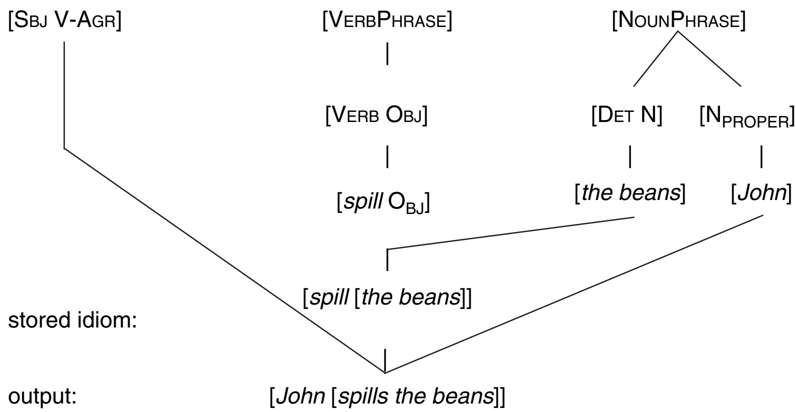

Figure 6.1 *Multiple inheritance network of* John spills the beans

subscribe to the idea that '[a]ny construction with unique idiosyncratic morphological, syntactic, lexical, semantic, pragmatic or discourse-functional properties must be represented as an independent node in the constructional network in order to capture a speaker's knowledge of their language' (Croft and Cruse 2004: 263). Nevertheless, in *John spills the beans* the idiom behaves like an ordinary VP, i.e. it follows the subject. Furthermore, the verb *spill* shows standard verb agreement and the object *the beans* is created like an ordinary NP. The taxonomic network in Figure 6.1 captures these facts by allowing specific constructions to inherit properties from more general, abstract constructions: the verb *spill*, for example, agrees with *John* because the construction inherits subject–verb agreement from the more general [SBJ V-AGR] construction. *The beans* and *John*, on the other hand, inherit the properties normally associated with NPs.

Researchers working within the Construction Grammar framework assume that all language-specific generalizations can be captured by multiple inheritance networks ('defaults') such as Figure 6.1 (cf. e.g. Croft and Cruse 2004: 262–5; Ginzburg and Sag 2000: 5–8; Goldberg 2003: 222–3). These inheritance networks allow actual expressions ('constructs') to be freely formed as long as the constructions they consist of do not conflict (Goldberg 2006: 22; more on this below). Consequently, '[c]onstructional approaches share with mainstream generative grammar the goal of accounting for the creative potential of language (Chomsky 1957, 1965). That is, it is clear that language is not a set of sentences that can be fixed in advance. Allowing constructions to combine freely as long as there are no conflicts, allows for the infinitely creative potential of language' (Goldberg 2006: 22).

Construction Grammar approaches thus provide a framework that not only accounts for both the core linguistic structures and all phenomena labelled peripheral by Principles-and-Parameters approaches, but also models the generative capacity of a speaker's mental grammar. It is for these reasons that I take Construction Grammar approaches to be more descriptively adequate than Principles-and-Parameters theories. Yet how explanatorily adequate is Construction Grammar?

Another central tenet held by a majority of proponents of Construction Grammar is that '[c]onstructions are understood to be learned on the basis of the input and general cognitive mechanisms' (Goldberg 2003: 219). This view contrasts sharply with the idea of an innate language faculty as advocated by Principles-and-Parameters theories. Usually, the main reason for the postulation of an innate language faculty depends on the 'poverty of stimulus' argument:

> It seems clear that many children acquire first and second languages quite successfully even though no special care is taken to teach them and no special attention is given to their progress. It also seems apparent that much of the actual speech observed consists of fragments and deviant expressions of a variety of sorts. Thus it seems that a child must have the ability to 'invent' a generative grammar … even though the primary linguistic data that he uses … may … be deficient in various respects. (Chomsky 1965: 200–1)

Chomsky thus assumes that input alone cannot explain why language acquisition is such a fast and uniform process, which means that children must be genetically predisposed to acquire languages (which in the end led him to postulate Universal Grammar).

Yet, this is, of course, an empirical issue that requires a closer look at the kind of input children actually receive as well as the structures they produce. Pullum and Scholz (2002), for example, have shown that the positive input children are exposed to is in fact far richer than assumed by mainstream generative grammar. Take, for example, the declarative sentence *The child that is$_1$ alone is$_2$ unhappy*, which has the corresponding yes-or-no question *Is$_2$ the child that is$_1$ alone unhappy?*, and not **Is$_1$ the child that alone is$_2$ unhappy?*. Principles-and-Parameters textbooks use these constructions to claim that structure-dependency must be part of Universal Grammar: '[c]hildren learning English probably never hear any sentences of this type' (Cook and Newson 1996: 13), yet they never simply front the first auxiliary *is$_1$*, but only the correct main clause auxiliary *is$_2$*. However, as Pullum and Scholz are able to prove, children receive positive input for auxiliary-initial sentences such as *Where's the other dolly that was in here?* by their caregivers (2002: 44). Furthermore, they also show that children can encounter positive evidence for other structures usually deemed absent from their input (such as plurals in noun-noun compounding auxiliary sequences or the correct use of anaphoric *one*; see Pullum and Scholz 2002: 24–36).

The primary linguistic input thus appears complex enough for children to build their mental grammars,[1] but how do they actually acquire language? Tomasello (2003, 2006) argues that language acquisition can in fact best be described by employing Construction Grammar models. As he notes, empirical studies on the linguistic output of children (see e.g. Diessel 2006; Diessel and Tomasello 2000; Lieven *et al.* 2003) show that children first (around the age of 14 months) employ holophrases, i.e. linguistic symbols such as *Birdie!* or *Lemme-see!*, which they treat as unanalysed chunks to express their intentions with respect to a specific scenario (Tomasello 2006: 23). In a next step, around the age of 18 months, children start tweaking utterance-level constructions

> to fit the current communicative situation. The basic ways they can do this are: (1) filling a new constituent into a slot in the item-based construction (as when *I wanna – –* and *ball* combine to make *I wanna ball*); (2) adding a new constituent onto the beginning or end of an utterance-level construction or expression (as when *Throw it* and *here* combine to make *Throw it here*); and (3) inserting a new constituent into the middle of an utterance-level construction or expression (the way a German child might insert *auch* ['too'] into a schema position where nothing had ever before appeared). (Tomasello 2003: 308–9)

Around the age of 18–20 months, so-called item-based constructions surface in which syntactic marking such as word order or inflectional morphology is used for the first time to signal the role of a participant in a scenario (e.g. *X hit Y*, *Y broken*, *put X in/on Y*). However, as Tomasello points out, while these constructions can have more than one open slot they are centred on a fixed substantive lexical item (thus only *hit* and not *beat* or *punch* might be used for a scenario where X beats Y; cf. Tomasello 2006: 24). Only later, at the age of about six, do children exhibit utterance-level constructions of adult-like abstractness (such as the abstract ditransitive or the passive construction; see Tomasello 2003: 316; 2006: 24).

What Tomasello has shown is that language learning is a gradual process in which atomic substantive constructions (i.e. holophrases) enter the lexicon first. Then more and more schematic constructions are acquired, resulting finally in a mental grammar that ranges from atomic substantive to complex schematic constructions (as illustrated in Table 6.1). According to Tomasello, there are four psycholinguistic processes that underlie language acquisition (2003: 295–305; 2006: 27–32):

[1] There are also researchers working within the Construction Grammar paradigm, notably Culicover and Jackendoff (2005), who believe that at least some aspects of language are genetic (cf. Jackendoff 2002: 69–103 for an overview of the arguments for this position, and Tomasello 2003: 284–90 for a critical view). Besides, note that even within the Principles-and-Parameters approach there have been considerable changes as to which purely linguistic features are innate (cf. Hauser, Chomsky and Fitch 2002, who claim that only recursion is part of the narrow faculty of language).

(1) **Intention reading and cultural learning**: in order to express intentions children must learn to read the intentions of others and know how these are linguistically encoded in a culture.

(2) Generalization principles: **schematization** (the replacement of substantive items by schematic slots) and **analogy** (by comparing two structures *X is Y-ing the Z*, e.g. *the cat is killing the mouse*, and *V is Y-ing the W*, e.g. *the boy is killing the bird*, children can generalize construction-specific semantic roles such as 'killer' and 'victim'; furthermore, once *A is B-ing the C*, e.g. *Daddy is kissing Mommy*, is contrasted with the former two examples, even more general syntactic and semantic roles such as 'agent'–'patient' and 'transitive subject'–'transitive object' emerge via analogy).

(3) **Entrenchment** and **preemption**: schematization and analogy only apply to constructions which are entrenched, i.e. which have been encountered and used frequently. This statement, however, needs to be qualified in that the two generalization principles only affect constructions with a high type-frequency, i.e. structures that have been encountered with many different lexicalizations, all of which have something in common, i.e. are semantically related (see Bybee 1985, 1995; Croft and Cruse 2004: 308–13; Goldberg 2006: 98–101). As Goldberg puts it: 'a pattern is considered extendable by learners only if they have witnessed the pattern being extended' (2006: 99). If a construction has an exemplar with high token frequency, i.e. a particular lexicalization that is used over and over again, then this will only lead to the entrenchment of this particular substantive construction (cf. Croft and Cruse 2004: 292–5).

In addition to this schematization and analogy are restricted by preemption: if there are two different ways, Form X and Form Y, of expressing a message and a speaker uses Form X, the hearer will assume that 'there was a reason for that choice related to the speaker's specific communicative intention' (Tomasello 2003: 300). In other words, the hearer seeks to minimize constructional synonymy by functionally distinguishing the two constructions.

(4) **Functionally based distributional analysis**: based on previous experience, newly learned items can be used 'the way other "similar" items have been used in the past – with no direct experience' (Tomasello 2003: 301). Paradigmatic categories like noun and verb, for example, 'are formed through a process of functionally based distributional analysis in which concrete linguistic items (such as words or phrases) that serve the same communicative function in utterances and constructions over time are grouped together into a category' (Tomasello 2003: 301).

In addition to being descriptively adequate, the increasing research on first-language acquisition (Diessel 2006; Diessel and Tomasello 2000; Lieven *et al.* 2003; for an overview cf. also Goldberg 2006: 69–92) proves Construction

Grammar approaches as potentially explanatorily adequate. Moreover, the input-driven, usage-based processes of increasing schematization of constructions also appears to be an adequate model for second-language acquisition (see e.g. N. Ellis 2003; Haberzettl 2006).

In light of the above findings, I take Construction Grammar approaches to be the most descriptively and explanatorily adequate account of mental linguistic knowledge. Before moving on to the Construction Grammar analysis of preposition placement in British and Kenyan English, however, it is necessary to point out that at present there does not exist one single Construction Grammar theory. Instead there are many different constructionist approaches (for a concise overview, see Goldberg 2006: 205–26), the most prominent ones being Unification Construction Grammar (Fillmore 1999; Kay and Fillmore 1999; related to this are Head-Driven Phrase Structure Grammar (HPSG)-based Construction Grammars as outlined by Ginzburg and Sag 2000; Sag 1997), Cognitive Grammar (Langacker 2005), Radical Construction Grammar (Croft 2001) and Cognitive Construction Grammar (Goldberg 2003, 2006; Lakoff 1987). All these theories share the major assumptions outlined above: they are non-derivational syntactic theories which take all levels of grammatical description to involve a structured default[2] inheritance network of constructions (see Goldberg 2006: 215). Yet, while they all agree that idiosyncratic properties lead to the postulation of an independent construction, there is disagreement as to the role of the frequency of constructions in language use: on the one hand, Cognitive Grammar, Radical Construction Grammar and Cognitive Construction Grammar are so-called ‘usage-based models’, i.e. they advocate that the frequent use of a construction can lead to it being cognitively entrenched, even if its properties can be completely derived compositionally by the underlying subconstructions. In the ‘complete inheritance model’ of Unification Construction Grammar, on the other hand, only idiosyncratic properties justify the existence of a construction (for an overview of the discussion, see Croft and Cruise 2004: 276–8; Goldberg 2006: 213–17).

Considering the above discussion about language acquisition, it should be clear that usage-based approaches are psychologically more plausible (see Goldberg 2006: 215). Speakers construct their inheritance construction networks on the basis of an input that just consists of substantive elements. More abstract, general constructions only emerge via generalization processes such as schematization and analogy. However, while I take usage-based approaches to be correct with respect to their psychological plausibility, there is one assumption of the three models currently working within this

[2] Default means that a more general, schematic construction contributes all its information to a more specific one unless the latter construction contains specific information which overrides the more general one.

paradigm that I disagree with: Cognitive Grammar, Radical Construction Grammar and Cognitive Construction Grammar eschew formalization, relying on a 'notation developed for ease of exposition only' (Goldberg 2006: 215). The main reasons for this, according to Goldberg, are that

> [U]nification-based approaches are not sufficiently amenable to capturing detailed lexical semantic properties. (2006: 216).

> Use of a fixed set of features or tools for even formal aspects of constructions is at odds with Croft's (2001) position that grammatical categories and roles are not general across constructions but are only defined with respect to particular constructions. (2006: 216)

> To linguists unfamiliar with the formalisms, unification can appear dauntingly opaque and cumbersome. (2006: 217)

I find all of these three reasons somewhat unconvincing: I personally also prefer linguistic articles that are easy to read and do not require me to consult some reference work every two pages or so. Nevertheless, if linguistics is supposed to qualify as a science then what matters is that a theory correctly describes the data and contains no contradictory mechanisms which render its predictions vacuous. Only when these criteria are met can we start worrying about intuitively accessible presentations of a model. This is especially important for Construction Grammar theories, which postulate that a speaker's mental grammar consists of an incredibly large number of constructions that interact with each other in a complex inheritance network as exemplified by Figure 6.1 above. Only a precise, formalized model of such a network can ensure that it does not contain constructions whose combination is required to give a grammatical sentence but whose constraints conflict. As Pollard and Sag have correctly noted, 'as theories become more complicated and their empirical consequences less straightforwardly apparent, the need for formalization arises' (1994: 6).

The same line of reasoning leads me to reject the objection that grammatical categories and roles need to be defined individually for each construction. Psychologically this might turn out to be correct, but at present such an approach also runs the danger of leading to conflicting construction constraints within the full inheritance network of a language, which compromises the observational and descriptive adequacy of a grammar.

Finally, the present work is not concerned with lexical semantic properties, which allegedly cannot be captured by unification-based approaches. Even if this position was correct, I take it that any formal Unification Construction Grammar account can easily be complemented by the sort of informal semantic description advocated by Goldberg.

Considering all of the above arguments, for the description of preposition placement in British and Kenyan English I decided on a combination of complete-inheritance and usage-based models, which could be labelled Usage-Based HPSG Construction Grammar: first I will present the constructions that due to their idiosyncratic properties warrant the postulation of independent constructions in both usage-based and complete-inheritance models. The unification-based version of Construction Grammar employed for this will be an HPSG-based one (as outlined in Ginzburg and Sag 2000; Sag 1997). This analysis will result in a construction network framed in a fully-fledged grammar formalism that is both explicit and falsifiable and has widely been used for the description of a great number of phenomena in various languages. After this, I will show how this complete-inheritance network can be enriched by usage-based models via the addition of constructions which due to their significant type and token frequency must be considered entrenched in the mental grammars of British and Kenyan speakers.

From a conceptual point of view, such an approach is probably not likely to win any favours with either of the two Construction Grammar camps. However, I would argue that such an eclectic approach ensures that my analysis is much more falsifiable than a purely usage-based one. In addition to this, while I obviously present my results in a formalized top-down fashion, I think it should be no problem translating it into an informal Cognitive Grammar, Radical Construction Grammar or Cognitive Construction Grammar analysis. In other words, those readers who only believe in complete-inheritance models are free to only read section 6.2 and ignore the enriched usage-based account given in section 6.3. Usage-based Construction Grammarians, on the other hand, can skim-read section 6.2, ignoring most of the formalization issues. Instead they can focus on the constructions that I postulate exhibit idiosyncratic properties and warrant storage in the inheritance network. In section 6.3 they will then find the full usage-based network, which they will probably be most interested in.

6.2 Preposition placement: The minimum complete-inheritance construction network

HPSG is a non-derivational unification-based grammar first developed by Pollard and Sag (1987, 1994). In the following I shall give a short overview of those central concepts of HPSG particularly pertinent to the discussion of preposition placement below (especially its Construction Grammar versions; cf. Ginzburg and Sag 2000; Sag 1997). For a detailed introduction to HPSG, which is beyond the scope of the present work, the reader is referred to Ginzburg and Sag (2000: 17–60).

The fundamental units of HPSG are signs, which are conceived of as 'structured complexes of phonological, semantic, discourse, and phrase-structural information' (Pollard and Sag 1994: 14). Figure 6.2 illustrates this

$$
\begin{bmatrix}
\textit{word} & \\
\text{PHONOLOGY} & \langle \textit{rely} \rangle \\
\text{SYNSEM} & \begin{bmatrix} \textit{synsem} & \\ \text{LOC} & \begin{bmatrix} \textit{local} & \\ \text{CAT} & \begin{bmatrix} \textit{cat} & \\ \text{HEAD} & \begin{bmatrix}\textit{verb} & \\ \text{AUX} & - \end{bmatrix} \end{bmatrix} \\ & \begin{bmatrix} \text{SUBJ} & \langle \boxed{3} \rangle \\ \text{SPR} & \langle\ \rangle \\ \text{COMPS} & \langle \boxed{4} \rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{soa} & \\ \text{NUCL} & \begin{bmatrix} \textit{rely-rel}(\textit{ation}) & \\ \text{RELY-ER} & \boxed{1} \\ \text{RELY-EE} & \boxed{2}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix} \\
\text{ARG-STR} & \langle \boxed{3}\,\text{NP}_{\boxed{1}}, \boxed{4}\,\text{PP}\begin{bmatrix}\textit{on}, \text{NP}_{\boxed{2}}\end{bmatrix} \rangle
\end{bmatrix}
$$

Figure 6.2 *An HPSG lexicon entry for* rely *(modelled on Ginzburg and Sag 2000: 18)*

by giving a simplified[3] HPSG lexicon entry for the non-passive form of the verb *rely*.

The first thing to note is that HPSG employs an Attribute-Value Matrix (AVM; see Pollard and Sag 1994:20). A concise definition of AVMs is provided by Goldberg:

Each attribute can have at most one value. Attributes may be n-ary, or may be features structures themselves. Any pair of AVMs can be combined to license a particular expression, just in case there is no value conflict on any attribute. When two AVMs unify, they map onto a new AVM, which has the union of attributes and values of the two original AVMs. (2006: 216)

While the AVM in Figure 6.2 might appear 'opaque and cumbersome' (Goldberg 2006: 217), it basically is nothing but a formalization of the various pieces of semantic and syntactic information associated with the phonological string /rɪ'laɪ/: the word has two arguments, one of which has to be syntactically encoded as a PP headed by *on*. This information has been formalized as follows.

[3] Note that the SUBJ, SPR and COMPS lists are in fact first created from the ARG(UMENT)-ST(RUCTURE) values via the Argument Realization Principle (see Ginzburg and Sag 2000: 171). Furthermore, the COMPS representation in the figure is somewhat simplified: the PP SYNSEM object will actually need to identify a PP object whose HEAD-DAUGHTER has the PHONOLOGICAL value ⟨*on*⟩ and whose other DAUGHTER will be specified as NP $\boxed{2}$.

As Figure 6.2 shows, one attribute of the sign contains its phonological form ('PHONOLOGY'), which for the sake of readability is given orthographically (cf. Pollard and Sag 1994: 15). In Figure 6.2 the attribute PHONOLOGY thus has the value ⟨ *rely* ⟩. In addition to that, syntactic and semantic information affecting the subcategorization of a sign are provided under the attribute SYNSEM. SYNSEM consists of a set of nested attributes, including information on LOC(AL) features (for NON-LOC features, see below). LOC features contain an attribute indicating the HEAD of a sign (thus the value *verb* for *rely* in Figure 6.2), as well as values specifying the set of permissible SUBJ(ECT)s, SP(ECIFIE)Rs and COMP(LEMENT)S licensed by this head. Another attribute nested under LOC is CONT(ENT), which gives the syntactically relevant semantic relations. In Figure 6.2 CONT contains the information that the semantic NUC(LEUS), i.e. the syntactic HEAD of the sign, denotes a *state-of-affairs* which is a 'rely-relation' between a RELY-ER and someone or something relied on, the RELY-EE (these values can thus be thought of as item-specific semantic roles). Finally, due to tags indicating token identity, the values of the ARG(UMENT)-ST(RUCTURE) attribute ensure that the RELY-ER $\boxed{1}$ is associated with an NP in the SUBJect $\boxed{3}$ position. Furthermore the NP marked as $\boxed{2}$ contained in the PP complement of *rely* $\boxed{4}$, which must obligatorily be headed by the preposition *on*, is identified with the RELY-EE $\boxed{2}$ (for details, cf. the Argument Realization Principle in Ginzburg and Sag 2000: 23).

The reinterpretation of an HPSG sign like Figure 6.2 as a construction is fairly obvious since it can also be considered the formalization of a learned pairing of form and meaning. Moreover, due to the uniform AVM representation, phrases also have a construction-type shape within HPSG. Take, for example, the VP *rely on John*, which has the HPSG representation in Figure 6.3. Again, Figure 6.3 has been simplified to contain only attributes pertinent for the present discussion. The structure in Figure 6.3 is of the type *h*(*ea*)*d-comp*(*lement*)*-ph*(*rase*) and, as the SYNSEM value indicates, is a VP. The D(AUGH)T(E)RS attribute of this sign contains as its value the list of immediate constituents: in Figure 6.3, the VP thus has a first daughter identified as $\boxed{2}$ and a second one which is the PP *on John* (tagged as $\boxed{3}$). Besides this, the H(EAD)-D(AUGH)T(E)R attribute indicates the syntactic head of the phrase (in Figure 6.3 this is the verb *rely*, which is tagged as $\boxed{2}$).

The VP in Figure 6.3 obviously has the same form as the sign in Figure 6.2, indicating that HPSG phrases are also form–meaning pairings that can easily be reinterpreted as constructions. Nevertheless, for the purpose of further illustration, the structure of Figure 6.3 has been represented by the more common tree format, which some readers might be more familiar with, in Figure 6.4. As this figure shows, the HD-DTR is the verb *rely*, which makes the projected phrase a VP. In addition to that, since this phrase contains an appropriate PP complement for *rely*, the complement subcategorization requirements of *rely* are satisfied, which is marked by the fact that the

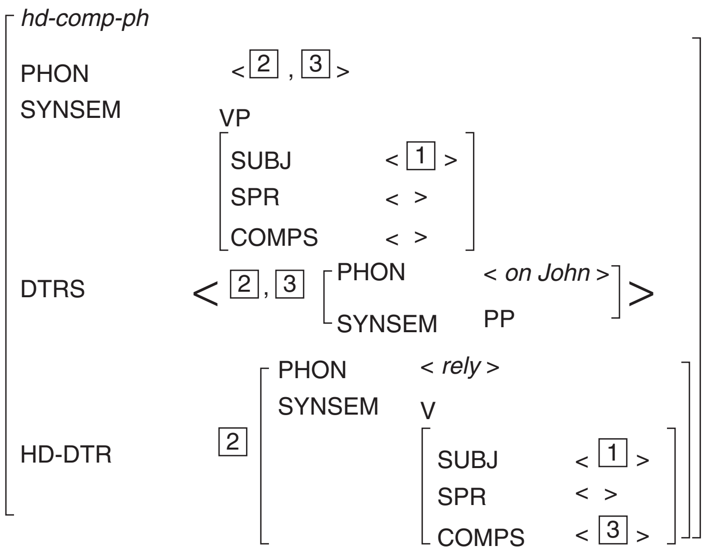

Figure 6.3 *The VP* rely on John *in HPSG (partly modelled on Ginzburg and Sag 2000: 30)*

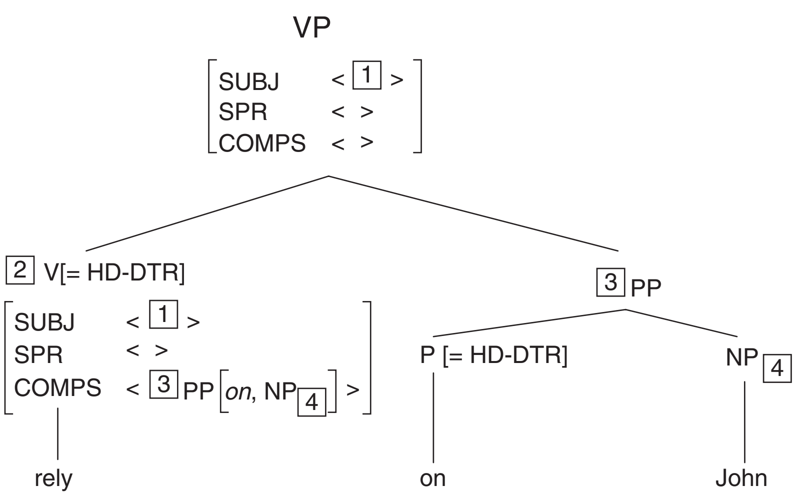

Figure 6.4 *The tree-notation of the HPSG VP* rely on John

entire VP has an empty COMPS list. All that still needs to be introduced is an eligible syntactic object $\boxed{1}$ in subject position.

In the complete-inheritance network of HPSG the VP in Figure 6.3 must not be stored in the lexicon since its properties derive from several independent constraints: the fact that a head daughter determines the type of

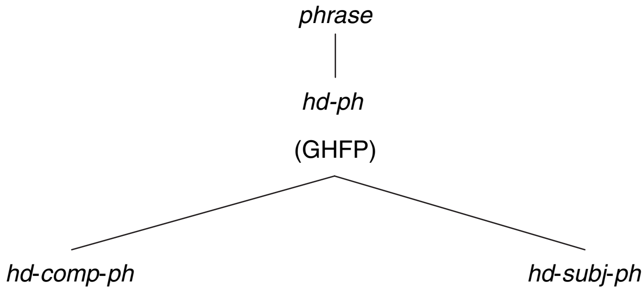

Figure 6.5 *The type hierarchy of selected headed phrases (taken from Ginzburg and Sag 2000: 32)*

a headed phrase is guaranteed by the Generalized Head Feature Principle (GHFP; by default the SYNSEM values of mother and head daughter of a phrase are identical; cf. Ginzburg and Sag 2000: 33). The properties of head-complement structures (e.g. that complement requirements which have been satisfied are deleted from the COMPS list of a phrase), on the other hand, are described by a constraint on the *h*(*ea*)*d-comp*(*lement*)*-ph*(*rase*) sign (see Ginzburg and Sag 2000: 34). The *h*(*ea*)*d-subj*(*ect*)*-ph*(*rase*) then handles phrases which contain the subject of a clause (cf. Ginzburg and Sag 2000: 34). The relationship between these signs and their constraints can be captured by an inheritance network: see Figure 6.5. One type of phrase is thus *h*(*eade*)*d-ph*(*rase*), which itself consists, among others (see Ginzburg and Sag 2000: 32 and below for the full network), of *head-complement-* and *head-subject-phrases*. In accordance with general properties of inheritance networks, the GHFP of the superordinate *headed phrase* applies by default to the subordinate *head-complement-* and *head-subject-phrases*, unless specific properties of these override this general constraint. One such specific property which overrides the GHFP is, for example, the deletion of COMPS features if an eligible complement is part of the phrase.

As pointed out above, the architecture of HPSG is well-suited for Construction Grammar approaches: both its fundamental types, words and phrases, are signs which can be seen as formalized form–meaning pairings. In addition to this, Construction Grammar versions of HPSG have extended the original multiple-inheritance network used for 'cross-classifying generalizations about words' (Ginzburg and Sag 2000: 5) to also capture 'generalizations about constructions in terms of cross-classifying type hierarchies' (Ginzburg and Sag 2000: 5). In Ginzburg and Sag's HPSG-based Construction Grammar, phrases are not only classified as to whether they are headed or not, 'but also relative to an independent dimension of "clausality"' (2000: 38). Figure 6.6 is a complex inheritance network with the two maximally general constructions HEADEDNESS and CLAUSALITY. For the present study, it is particularly the different CLAUSALITY constructions

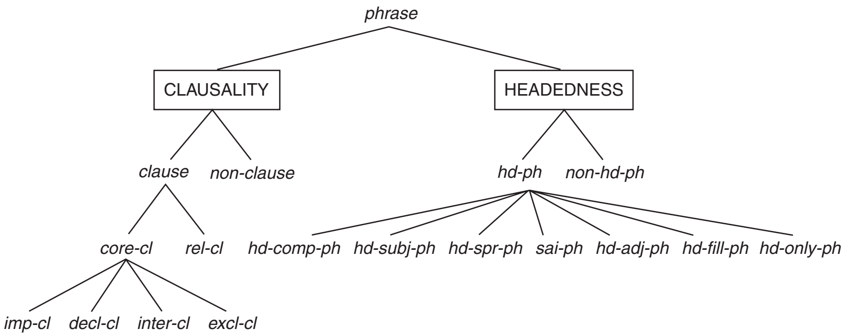

Figure 6.6 *HEADEDNESS and CLAUSALITY constructions (taken from Ginzburg and Sag 2000: 39)*

and the *h*(*ea*)*d-fill*(*er*)*-phrase* which will be of importance (for a fuller discussion of Figure 6.6, see Ginzburg and Sag 2000: 38–60).

Next I shall show how variable preposition placement can be captured in an HPSG Construction Grammar inheritance network, discussing potential modifications required by the British and Kenyan English empirical data presented in the previous two chapters. Lastly, I shall then apply the same procedure to categorically stranding clause types.

6.2.1 *Variable preposition placement*

As will be remembered, variable preposition placement in English surfaces in preposed/topicalized ((6.2a), (6.3a)), interrogative ((6.2b), (6.3b)), exclamative ((6.2c), (6.3c)) and relative clauses ((6.2d), (6.3d)):

(6.2)	a.	**[Stranding]**$_i$ I've heard **of**$_i$.	[preposing]
	b.	**[What]**$_i$ is he talking **about**$_i$?	[interrogative]
	c.	**[What a great topic]**$_i$ he talked **about**$_i$!	[exclamative]
	d.	the structure [**[which]**$_i$ he talked **about**$_i$].	[*wh*-relative]
(6.3)	a.	**[Of stranding]**$_i$ I've heard$_i$.	[preposing]
	b.	**[About what]**$_i$ is he talking$_i$?	[interrogative]
	c.	**[About what a great topic]**$_i$ he talked$_i$!	[exclamative]
	d.	the structure [**[about which]**$_i$ he talked$_i$].	[*wh*-relative]

As a quick look at Figure 6.6 reveals, *top*(*icalized*)*-cl*(*auses*) are absent from the list of clause types. They are, however, included as a type of *core-cl*(*ause*) in the full type hierarchy provided by Ginzburg and Sag in their Appendix (2000: 262). Furthermore, as mentioned in section 3.1.2.1, topicalized clauses differ from the other three optional preposition placement clauses in that they contain no *wh*-element. As we will see, this has important consequences for the HPSG treatment of preposition placement in topicalized clauses.

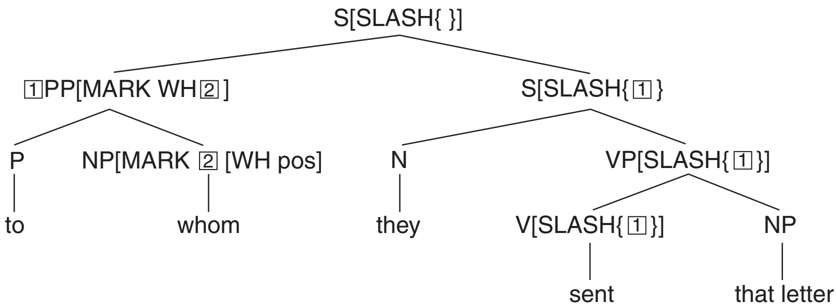

Figure 6.7 *Pied-piping in HPSG (modelled on van Eynde 2004: 315)*

Besides this, *rel*(*ative*)-*cl*(*auses*) differ from the other clauses because they function as modifiers and are always syntactically embedded, which leads to them being treated differently to the so-called *core-cl*(*auses*) such as *inter*(*rogative*)-*cl*(*auses*) or *excl*(*amative*)-*cl*(*auses*) (see Figure 6.6). For the treatment of pied-piping and stranding in HPSG this, however, has no consequence. Preposition placement in HPSG (following van Eynde 2004) can therefore be illustrated by using relative clause examples.

In HPSG-based accounts there are two general constructions that handle preposition pied-piping and stranding. Take, for example, the pied-piped relative clause in (6.4), which in HPSG has the structure given in Figure 6.7.

(6.4) the man [to whom they sent that letter]

A basic constraint on *wh*-questions, *wh*-exclamatives or *wh*-relative clauses in English (for topicalized clauses, see below) is the requirement that the non-head daughter (i.e. the position that corresponds to SpecC in the Principles-and-Parameters frameworks) carries a [WH pos] feature (cf. van Eynde 2004). In pied-piped cases such as (6.3), where the lexical [WH pos] feature of a *wh*-word is embedded in a prepositional phrase, the grammar must therefore provide a mechanism that ensures the percolation of that feature to the entire PP node. Figure 6.8 gives van Eynde's (2004) formalism for pied-piping in HPSG.

The construction in Figure 6.8 is of the type *hd-comp-ph* (cf. Figure 6.3 above). The HEAD-DTR is a preposition (as indicated by van Eynde in the SYNSEM features of the phrase), and the complement is marked as [WH]. The co-indexation (i.e. $\boxed{2}$) of the WH-features in Figure 6.8 ensures that the WH-feature value of the non-head daughter (i.e. the *wh*-word) percolates/is inherited by the entire prepositional phrase. As Figure 6.8 shows, the construction makes no reference to a specific clause type (e.g. relative or interrogative constructions). Thus it is implied that all cases of preposition pied-piping with *wh*-questions, *wh*-exclamatives or *wh*-relative clauses in English can be captured by a single constraint.

$$\left[\begin{array}{l} \textit{hd-comp-ph} \\ \text{SYNSEM} \mid \text{LOCAL} \mid \text{CAT} \left[\begin{array}{ll} \text{HEAD} & \textit{prep} \\ \text{MARKING} \mid \text{WH} & \boxed{2} \end{array}\right] \\ \text{DTRS} \left\langle \boxed{1}, \left[\text{SYNSEM} \mid \text{LOCAL} \mid \text{CAT} \mid \text{MARKING} \mid \text{WH} \ \boxed{2}\right] \right\rangle \\ \text{HEAD-DTR} \ \boxed{1} \left[\begin{array}{ll} \textit{word} & \\ \text{PHON} & \textit{list}\left(\textit{form}\right) \end{array}\right] \end{array}\right]$$

Figure 6.8 *Wh-pied-piping constraint in HPSG (van Eynde 2004: 329)*

Moving on to preposition-stranding, it is important to understand how HPSG models displacement phenomena: instead of assuming movement of a *wh*-phrase, filler–gap dependencies are handled by the percolation of SLASH features. If a lexically specified argument such as the receiver PP in (6.4) is not in its canonical complement position, a non-empty SLASH feature, a NON-LOC(AL) feature, will indicate that a COMPS element is missing locally (for the precise technicalities, cf. Ginzburg and Sag 2000: 167–71; van Eynde 2004: 315–17). In line with the Generalized Head Feature Principle, SLASH features percolate up the tree until an adequate filler phrase is encountered. A constraint on the *h*(*ea*)*d-fill*(*er*)*-ph*(*rase*) then ensures that the SLASH feature will be cancelled on the mother node if the non-head daughter qualifies as an appropriate filler for the missing COMPS object of the head daughter (for details, see Ginzburg and Sag 2000: 174). In Figure 6.7, *to whom* is obviously an appropriate filler for the missing COMPS gap (as indicated by the co-indexed feature tag $\boxed{1}$), which results in the SLASH feature being cancelled on the mother node S.

An important repercussion of the SLASH mechanism just outlined is that with respect to extraction pied-piping and stranding are handled alike, since in both cases an argument is not locally available to cancel a head's COMPS feature. The only difference is that in order to license stranded structures like (6.5a), the grammar must allow SLASH features in the Argument-Structure of prepositions:[4]

(6.5) a. $\boxed{1}$ **[who(m)]** they sent the letter **to[SLASH {$\boxed{1}$}]**
b. $\boxed{1}$ **[to whom]** they **sent[SLASH {$\boxed{1}$}]** that letter

[4] Since the clauses in question are normally verbal projections, the percolation of the SLASH feature must be mediated by the main verb. Technically this is ensured by the SLASH-Amalgamation Constraint (see Ginzburg and Sag 2000: 169), which states that words will always amalgamate 'the SLASH values of the members of their ARG-ST list' (Ginzburg and Sag 2000: 168).

Both the pied-piped (6.5b) and the stranded clause (6.5a) are thus instances of head–filler structures, which also covers object-extracted instances like *who*(*m*) *he loves*. The only special construction for variable preposition placement which has to be postulated in HPSG is van Eynde's *wh-pied-piping* one. The constraints of this construction (Figure 6.8) ensure that the non-head daughter in *wh*-questions, *wh*-exclamatives or *wh*-relative clauses is *wh*-marked.

What has to be noted is that according to the analysis just outlined an additional pied-piping construction (Figure 6.8) is only needed for *wh*-questions, *wh*-exclamatives and *wh*-relative clauses. Due to the fact that topicalized clauses do not contain a *wh*-item, preposition-stranding and pied-piping in these clauses are both covered by a combination of SLASH features and the *head-filler-phrase* construction:

(6.6) a. $\boxed{1}$ **[Stranding]** I've heard **of[SLASH {$\boxed{1}$}]**.
b. $\boxed{1}$ **[Of stranding]** I've **heard[SLASH {$\boxed{1}$}]**.

In (6.6a) the NP complement $\boxed{1}$ of the preposition *of* is missing locally and therefore SLASH-ed. Since the NP $\boxed{1}$*stranding* qualifies as an adequate filler this SLASH feature can be cancelled off. Similarly in (6.6b) the PP complement $\boxed{1}$ of *heard* is SLASH-ed, and the SLASH feature is cancelled off by the PP filler $\boxed{1}$ *of stranding*.

In order to see how the above analysis fits into the construction network of Figure 6.6, the constraints required to capture preposition placement in English have been incorporated into this network in Figure 6.9. As this figure shows, the types *wh-interrogative-clause*, *wh-exclamative-clause*, *topicalized-clause* and *fin*(*ite*)*-wh-filler-relative-clause* are all of the type *head-filler-phrase* (though modifications of this will be necessary for the relative clauses; cf. below). This means that in all of them an overt filler cancels off a SLASH-ed COMPS feature. The reason why SLASH[5] is given as an independent construction is that it will also be employed for the categorically stranding clauses, which, as will be seen below, do not employ an overt filler. Furthermore, the *wh-pied-piping* construction only applies to *wh*-clauses and not to constructions of the type *topicalized-clause*.

A few more comments on Figure 6.9 seem necessary. First of all, in Ginzburg and Sag's analysis, *subject-wh-interrogative clauses* (e.g. *Who cares about preposition stranding?*) are also of the type *head-filler-phrase* (2000: 236–40). Consequently, as expected from a complete-inheritance approach, the *head-filler-phrase* need only apply to the general *wh-interrogative-clause* construction. All the more specific constructions, such as *n*(*on*)*-su*(*bject*)*-wh-interrogative-clause*, will inherit the constraints of the

[5] In Figure 6.9 SLASH is merely a short-hand notation for Ginzburg and Sag's more complex (and technically more precise) *gap-ss* constraint (2000: 167–74).

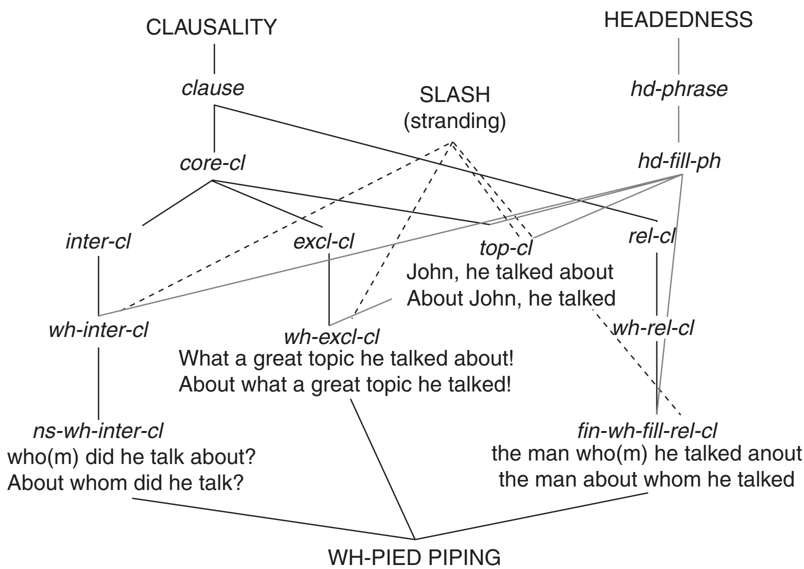

Figure 6.9 *Minimal HPSG preposition placement construction network (preliminary version based on Ginzburg and Sag 2000; Sag 1997)*

head-filler-phrase by default. Secondly, the *wh-pied-piping* construction in Figure 6.9 seems limited to the *finite-wh-filler-relative-clause* construction but this should only be seen as an intermediate result. As will be remembered, relative clauses exhibit the most complex interaction of categorical constraints and will therefore be investigated in more detail below.

Finally, it must be emphasized again that the SLASH construction is not limited to preposition-stranding (which in Figure 6.9 is indicated by putting 'stranding' in parentheses). Instead, it also licenses object gaps:

(6.7) a. $\boxed{1}$ **[Stranding]** he **likes[SLASH {**$\boxed{1}$**}]**. [preposing]
b. $\boxed{1}$ **[What]** does he **like[SLASH {**$\boxed{1}$**}]**? [interrogative]
c. $\boxed{1}$ **[What a boring topic]** this **is[SLASH {**$\boxed{1}$**}]**! [exclamative]
d. a structure [$\boxed{1}$ **[which]** he **hates[SLASH {**$\boxed{1}$**}]**]. [*wh*-relative]

For a complete-inheritance model the possibility of such a generalization is obviously a welcome result. Cross-linguistically, however, this raises the question why there are languages such as German in which preposition-stranding is ungrammatical:

(6.8) a. $\boxed{1}$ **[Mit wem]** hat er **gesprochen[SLASH {**$\boxed{1}$**}]**?
'To whom did he talk'
b. *$\boxed{1}$ **[Wem]** hat er **mit[SLASH {**$\boxed{1}$**}]** gesprochen?
'Who(m) did he talk to'

The simplest way to capture this fact in a complete-inheritance model is to ban preposition-stranding in such languages (cf. Culicover and Jackendoff 2005: 319). In other words the grammar must contain a constraint of the sort *P[SLASH {$\boxed{1}$}]. Economical as such an approach might be, a psychologically more adequate usage-based account would take a positive data approach, i.e. stranded prepositions are only licensed in languages in which speakers receive positive input for such structures. This issue will be returned to in section 6.3.

The construction network in Figure 6.9 captures the general distribution of preposition-stranding and pied-piping across those clause types which allow for variable preposition placement. It does not incorporate the fact that pied-piping has been claimed to be generally more formal than stranding (cf. section 3.3). Since the basic definition of signs in HPSG explicitly includes discourse information (cf. the quote from Pollard and Sag 1994: 14 above), it should, however, be possible to model formality effects in HPSG. In fact this has already been suggested by Wilcock (1999). He argues that formality can be implemented in HPSG by introducing the new attribute REG(I)ST(E)R as part of CONTEXT (an attribute that 'contains linguistic information that bears on certain context-dependent aspects of semantic interpretation' (Pollard and Sag 1994: 3)). According to Wilcock, the REGISTER attribute then has the two possible values *formal* and *informal* (1999: 382). Explicitly addressing preposition placement, he goes on to claim that formality effects can by captured by either clausal or lexical constraints.

Before discussing Wilcock's proposal in more detail, it should be remembered that the corpus studies in chapter 4 showed that the variables CLAUSE TYPE and FORMALITY exhibit interaction effects in both varieties:

- in British and Kenyan English free relative clauses, main questions and embedded interrogatives strongly favour stranding regardless of the level of formality;
- in British English *wh*-relative clauses, stranding is strongly favoured in informal contexts and pied-piping is strongly associated with formal text types, while
- in Kenyan English pied-piping is favoured with *wh*-relative clauses in all contexts, with informal text types only decreasing the strength of this preference.

One consequence of these results is that for both varieties it is not possible to simply employ a feature *informal* that is associated with preposition-stranding in general. First of all, such a constraint could not be introduced to the SLASH construction in Figure 6.9 simply for the technical reason that it covers various kinds of object extraction as well (cf. 6.7), which are not limited to informal contexts. Besides this, even if a *preposition stranding* P[SLASH {$\boxed{1}$}] subtype of the SLASH construction is postulated, this cannot have an informal REGISTER value, since stranded prepositions commonly occur in formal contexts in free relatives, main questions and embedded interrogatives. Instead, the locus of formality effects in both varieties are *wh*-relative clauses. It will therefore be necessary to take a closer look at the analysis of relative clauses in HPSG. In addition to this, since *that*- and Ø-relative clauses will be discussed as well, the next chapter can also be seen as paving the way for the analysis of the categorically stranding clauses in section 6.2.3.

6.2.2 *English relative clauses and Construction Grammar*

Working with an HPSG-based Construction Grammar approach, Sag assumes a network with six types of restrictive relative clause constructions (1997: 464, 473). As Figure 6.10 illustrates, Sag postulates three types of *wh-relative-clause* constructions:

- the *wh-subject-relative-clause* construction (e.g. *the man who left*),
- the *finite-wh-filler-relative-clause* construction (e.g. *the man who they like*),
- and the *non-finite-wh-filler-relative-clause* construction (e.g. *on which to depend*),

which all share the constraint that the non-head daughter is a *wh*-relativizer (i.e. an element with a REL-feature, thus the formalization NON-HD-DTRS ⟨REL {$\boxed{1}$} ⟩) that must be identified with the NP$\boxed{1}$ phrase which is modified by the relative clause (this is achieved via the additional constraint [HEAD [MOD NP$\boxed{1}$]]; cf. Sag 1997: 451).

Furthermore, he assumes the existence of three *non-wh-relative-clause* constructions:

- the *bare-relative-clause* construction (e.g. *the man Sandy likes*),
- the *simple-non-finite-relative-clause* construction (e.g. *the man to visit*), and
- the *reduced-relative-clause* construction (e.g. *the man standing here*).

These three constructions share the following constraint (taken from Sag 1997: 468):

(6.9) $$\textit{non-wh-rel-cl} \Rightarrow \begin{bmatrix} \text{HEAD} & [\text{MOD N}_{\boxed{1}}] \\ \text{SLASH} & \{\ \} \\ \text{HD-DTR} & [\text{SLASH } \{\text{NP}_{\boxed{1}}\}] \end{bmatrix}$$

As can be seen in Figure 6.10, none of the non-*wh*-relative clauses is analysed by Sag as being of the type *head-filler-phrase*. Yet, in both *bare-relative-clause* and *simple-non-finite-relative-clause* constructions an argument is missing locally and therefore SLASH-ed (cf. *the man* [*Sandy likes*[SLASH {$\boxed{1}$}]] and *the man* [*to visit*[SLASH {$\boxed{1}$}]], respectively). In order to guarantee

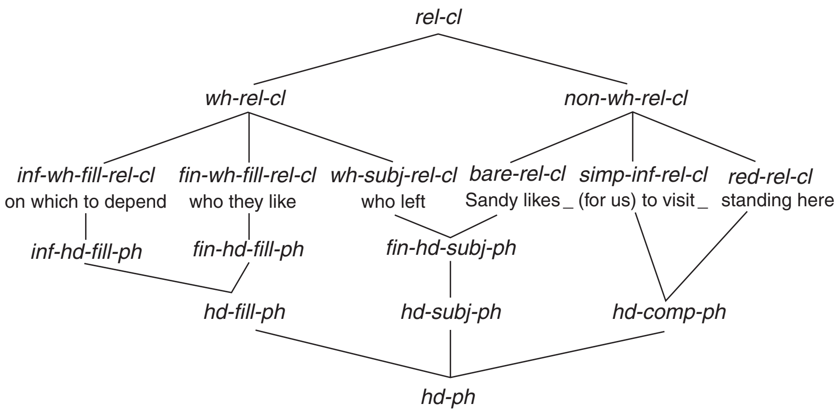

Figure 6.10 *HPSG relative clause construction network (adapted from Sag* 1997*: 464, 473)*

the grammaticality of these structures, the constraint in (6.9) ensures that this SLASH feature is cancelled off at the top of a non-*wh*-relative clause. Since these clauses do not employ an overt *wh*-filler, it follows that the *wh-pied-piping* construction does not apply, leading to categorical preposition-stranding in bare relative clauses (*the man she talked to*) and simple non-*finite* relative clauses (*the man to talk to*). Furthermore, the SLASH-ed element in (6.9) is indexed with the head noun of the antecedent NP, due to which the gap is referentially linked to this noun (cf. Sag 1997: 468). Finally, the [MOD N′$\boxed{1}$] requires that non-*wh*-relatives combine with the nominal head before this takes its required determiner (leading to a structure like [*the*] [*man* [*Sandy likes*]]; this accounts, for example, for the fact that Ø-relative clauses must precede *wh*-relative clauses, since the latter combine with the full antecedent NP; cf. Sag 1997: 465–8).

Figure 6.10 shows that the construction network involves subconstructions of the *head-filler-phrase* and *head-subject-phrase*: Sag, postulates, for example, that there is a *fin*(*ite*)-*head-subject-phrase* which requires verbal heads to be finite (1997: 441), while in Ginzburg and Sag such a constraint on the finiteness of heads can also appear on clause constructions (2000: 43). This, however, is a technical detail which is immaterial for the present discussion since subject relative clauses are not the focus of this study. The only point to note is that in Sag's analysis neither the *wh-subject-relative-clause* construction nor the *bare-relative-clause* construction involves an extracted filler. Instead the *wh*-phrase in *who left* in Figure 6.10 is interpreted as an argument in subject position just like *Sandy* in *Sandy likes*.

In addition to this, in Figure 6.10 the *head-filler-phrase* has two subconstructions: *inf*(*inite*)-*h*(*ea*)*d-fill*(*er*)-*ph*(*rase*) and *fin*(*ite*)-*h*(*ea*)*d-fill*(*er*)-*ph*(*rase*)). Sag motivates this distinction by pointing out that finite head-filler structures always require an overt subject (cf. **These bagels, Ø likes* or **the baker* [[*whose bagels*] *Ø likes*]; 1997: 454). In contrast to this, non-finite head-filler clauses do not allow overt subjects (cf. **the baker* [[*in whom*]] (*for*) *you to place your trust* or **I wonder* [[*in whom*] (*for*) *them to place their trust*]; Sag 1997: 461).

Turning to the specific relative clause constructions in Figure 6.10, it turns out that Sag's classification incorporates one type of obligatory pied-piped preposition construction: the *infinite-wh-filler-relative-clause* construction. All other properties of pied-piping and stranding are assumed to follow from the interaction of the above set of relative clause constructions and the general constructions licensing preposition placement discussed above. The postulation of the *infinite-wh-filler-relative-clause* construction, which is of the type *infinite-head-filler-phrase* and *wh-relative-clause*, is justified by the idiosyncratic obligatory pied-piping requirement of these clauses (cf. section 3.1):

(6.10) a. the man on whom to rely
b. *the man whom to rely on
(6.11) a. I wonder on who(m) to rely
b. I wonder who(m) to rely on

While non-finite *wh*-interrogative clauses allow both stranding and pied-piping (see (6.11)), preposition-stranding is prohibited in non-finite *wh*-relative clauses (cf. (6.10b); also Hoffmann 2005: 263). Sag captures this property by a constraint which requires the filler in *infinite-wh-filler-relative-clause* constructions to be of the type PP (i.e. [NON-HD-DTRS <PP>]; Sag 1997: 462).

Finite-wh-filler-relative-clause constructions are of the type *finite-head-filler-phrase* and *wh-relative-clause*. Apart from the constraints that they inherit from these two more general constructions, *finite-wh-filler-relative-clauses* have the additional restriction that their filler must either be an NP or a PP ([NON-HD-DTRS ⟨[HEAD *noun* v *prep*⟩]; Sag 1997: 454].[6] As pointed out in the preceding section, *finite-wh-filler-relative-clause* constructions are the ones to which both stranding (via the SLASH construction) and pied-piping (via the *wh-pied-piping* construction) can apply. It is therefore interesting to note that Sag analyses relative *that* as a *wh*-relative word (i.e. carrying a REL-feature; Sag 1997: 463). Accordingly, he considers *that*-relative clauses instantiations of the *finite-wh-filler-relative-clause* constructions. This, however, means that *that*-relative clauses should also allow pied-piping, since as *finite-wh-filler-relative-clause* constructions they are eligible for the *wh-pied-piping* construction. Yet, structures like **the man to that he talked* are clearly ungrammatical.

[6] This is necessary since other *finite-head-filler-phrase* constructions are less restrictive with respect to fillers they license e.g. questions also allow AdjP fillers (cf. *How cool am I?*; Sag 1997: 454).

Sag explains the ungrammaticality of pied-piping with *that* as the result of case mismatch and/or register restrictions: for him, *that* in Present-day English is a relative pronoun which carries nominative case and is limited to informal registers just like *who*. Prepositions, however, require their complements to carry oblique case. Consequently, the impossibility of P + *that* would be a case assignment violation similar to pied-piping with nominative *who*. Besides this, according to his account, both *who* and *that* are informal and pied-piping is formal, which might also lead to obligatory stranding with *that* and *who* (cf. 1997: 463). While this explanation also accounts for the fact that *that* cannot occur in non-finite relative clauses, where pied- piping is obligatory (Sag thus formalizes an idea already advocated in Van der Auwera 1985), the evidence from the experimental studies (chapter 5) clearly indicates that such an analysis cannot be entertained for either British or Kenyan English.

In the first Magnitude Estimation experiment (section 5.1), pied-piping with *that* and with Ø was judged as bad as the hard grammatical constraints such as word order violations. The second experiment (section 5.2) then contrasted P + *who* with P + *that* and P + Ø structures. As the results showed, pied-piping with *who* was judged similar to case mismatch effects but was still judged significantly better than pied-piping with *that* or Ø. For both British and Kenyan English speakers pied-piping with the latter two relativizers again yielded acceptability scores similar to the worst ungrammatical fillers (i.e. subject contact clauses). All of this proves that *that* is not on a par with *who* in either British or Kenyan English. Instead, in both varieties *that* behaves like Ø with respect to preposition placement.

The empirical data therefore suggests that *that*-relative clauses are more similar to *non-wh-relative-clauses* (i.e. relative clauses without an overt relative *wh*-pronoun), a conclusion which is supported by various variationist studies on factors influencing the choice of relativizer: both *that* and Ø are restricted to restrictive relative clauses and both are preferred in more informal contexts (see e.g. Ball 1996; Guy and Bayley 1995). Consequently, it seems empirically more adequate to treat *that*-relative clauses as a special type of *non-wh-relative-clause* with the following properties:

(6.12) *ns-that-fin-rel-cl* ⇒ [HEAD-DTR *that*]

(6.13) *ns-that-fin-rel-cl* ⇒ $\begin{bmatrix} \text{HEAD} & [\text{MOD NP}_{\boxed{1}}] \\ \text{SLASH} & \{\ \} \\ \text{HD-DTR} & [\text{SLASH}\ \{\text{NP}_{\boxed{1}}\}] \end{bmatrix}$

The first constraint of the *n(on)s(ubject)-that-finite-relative-clause* construction in (6.12) ensures the presence of a *that* complementizer, which like all complementizers functions as the head of the (relative) clause CP (cf. Sag 1997: 457). Due to the SLASH Amalgamation Constraint (Ginzburg and Sag 2000: 169), the head of the CP will inherit the SLASH features of its argument, i.e. the VP. This ensures that the information about the relativized gap is not lost. In a next step, the constraint in (6.13) ensures that the SLASH feature is not percolated beyond the relative clause, but is bound off. Basically, (6.13) is similar to the constraint on *non-wh-relative-clauses* (cf. Sag 1997: 468), with the only exception that *that*-relative clauses modify NPs and not N′. Consequently, *nonsubject-that-finite-relative-clauses* are not instantiations of a *head-filler-phrase* but inherit properties from the *finite-head-subject-phrase* (just like *bare-relative-clauses*). Finally, since *nonsubject-that-finite-relative-clauses* do not contain a *wh*-element, they are not eligible for the *wh-pied-piping* construction, which causes obligatory stranding with a *that*-relativizer.

Note that treating *that* as a finite complementizer in relative clauses also helps to explain why preposition-stranding never extended to non-finite *wh*-relative clauses: since the finite complementizer *that* never appeared in non-finite relative clauses, these structures lacked an overt relativizer + $P_{stranded}$ model. As a result, *wh*-relativizers retained their historically obligatorily pied-piping constraint in non-finite *wh*-relative clauses (see e.g. Allen 1980: 292; Fischer *et al.* 2000: 59).

Such an approach obviously also has repercussions for the analysis of *wh-subject-relative-clause* constructions. If *that* is not a regular *wh*-relativizer then sentences like (6.14) (taken from Huddleston, Pullum and Peterson 2002: 1047) cannot be captured by the *wh-subject-relative-clause* construction:

(6.14) I want a car that is safe.

While an analysis of subject relative clauses is beyond the scope of the present study, it should be noted that an extra *subject-that-finite-relative-clause* construction seems warranted. This construction would need to share certain properties with the *nonsubject-that-finite-relative-clause* construction (such as (6.12), i.e. the presence of a *that* HEAD-DTR and a mechanism that identifies the relative-clause-external antecedent as the correct logical subject of the relative clause). Figure 6.11 should be seen as an attempt to capture these features in HPSG-based Construction Grammar.

Figure 6.11 assumes that *subject-that-finite-relative-clause* constructions also modify NPs and that in this particular case *that* functions as a complementizer with pronominal characteristics: its CONTENT is identified with the relative–clause-external antecedent NP but it nevertheless functions as a complementizer. In addition, the construction in Figure 6.11 ensures that *that* mediates the identification of the subject of the relative and the antecedent NP. Finally, the construction combines properties of the

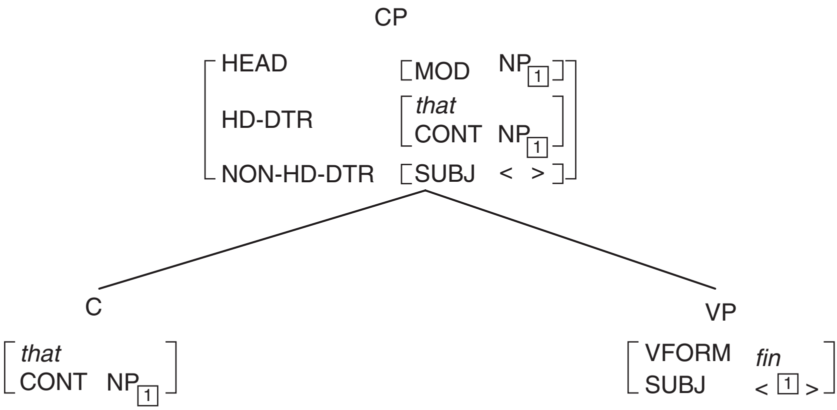

Figure 6.11 *A possible entry for the* subject-that-finite-relative-clause *construction*

head- complement-phrase(with the complementizer *that* functioning as head) as well as the *head-subject-phrase* (with *that* instantiating the SUBJ value of its complement), but obviously also overriding them (by combining the two). This analysis thus tries to account for the pronominal as well as complementizer features of relative *that*.

The constructions outlined above suffice to generate all of the relative clauses observed in British and Kenyan English. What still needs to be addressed is the formalisation of the level of formality. As mentioned above, Wilcock (1999: 382–4) suggests two possible ways in which formality can be incorporated into HPSG: either as clausal or lexical constraints. I will illustrate those two options in turn.

The first option Wilcock offers is to associate the attribute REG(I)ST(E)R with clause types. He explicitly mentions pied-piping in relative clauses and discusses the possibility of introducing a subtype of the *finite-wh-filler-relative clause* construction (1999: 383).[7] This *finite-wh-pp-filler-relative-clause* construction would be subject to the following constraint (taken from Wilcock 1999: 383):

(6.15) $$\textit{fin-wh-pp-fill-rel-cl} \Rightarrow \left[\text{FILLER-DTR} \begin{bmatrix} \text{HEAD} & \textit{prep} \\ \text{REGSTR} & \textit{formal} \end{bmatrix}\right]$$

As (6.15) shows, this construction requires a PP filler and the entire clause is marked as formal in its REGISTER attribute.

[7] Actually he also proposes a second construction for finite *wh*-relative clauses with an NP-filler *fin-wh-np-fill-rel-cl* (Wilcock 1999: 383). Since this construction has no additional properties I have omitted it. I take NP-filler cases to be covered by Sag's *finite-wh-filler-relative-clause* constructions. For the following analysis this, however, is of no consequence.

The second option put forward by Wilcock is to encode the REGSTR constraints on lexical subtypes of prepositions. Take the following set of abstract preposition constructions in (6.16) (adapted from Wilcock 1999: 384):

(6.16) a) *rel-prep* b) *que-prep* c) *slash-prep*

$$\begin{bmatrix} \text{HEAD} & \textit{prep} \\ \text{QUE} & \{\ \} \\ \text{REL} & \{\boxed{1}\} \\ \text{SLASH} & \{\ \} \\ \text{REGSTR} & \textit{formal} \end{bmatrix} \quad \begin{bmatrix} \text{HEAD} & \textit{prep} \\ \text{QUE} & \{\boxed{1}\} \\ \text{REL} & \{\ \} \\ \text{SLASH} & \{\ \} \\ \text{REGSTR} & \textit{formal} \end{bmatrix} \quad \begin{bmatrix} \text{HEAD} & \textit{prep} \\ \text{QUE} & \{\ \} \\ \text{REL} & \{\ \} \\ \text{SLASH} & \{\boxed{1}\} \\ \text{REGSTR} & \textit{informal} \end{bmatrix}$$

Construction (6.16a) encodes the fact that pied-piped prepositions with a *wh*-relativizer as their complement (thus the non-empty REL {$\boxed{1}$} feature) are marked as formal. The same mechanism is employed to mark pied-piped prepositions in questions ((6.16b); cf. the non-empty QUE {$\boxed{1}$} feature).[8] Finally, stranded prepositions (cf. the non-empty SLASH value in (6.16c)) are identified as informal by the value of their REGISTER attribute.

While Wilcock leaves open which formalization is the correct one, the corpus data for British and Kenyan English clearly show that lexical constraints such as (6.16b) and (6.16c) are empirically incorrect: in both varieties stranding is preferred with free relatives, main questions and embedded interrogatives regardless of the level of formality. The pied-piped construction in (6.16b) is only employed with adjunct PPs that resist stranding, and is not motivated by the level of formality for these clause types. This leaves only relative clauses, which as the empirical data proved, were the only clausal context affected by the level of formality. The question now is whether this effect should be captured in a lexical rule like (6.16a) or via a clausal construction such as (6.15).

I would argue that the British English data especially call for a clause-based explanation since the results from the corpus study showed that with finite *wh*-relative clauses pied-piping is strongly favoured in formal contexts, while in informal contexts stranding is strongly preferred. The lexical construction in (6.16a) is able to model the effect of formal pied-piped relative clauses, but no such explanation is possible for the stranded *wh*-data: if a preposition is pied-piped in a relative clause it occurs clause-initially, and in accordance with the *wh-pied-piping* constraint, will contain a WH-marker which will additionally be specified as a relative pronoun (by its REL value). In pied-piped relative clauses the preposition thus has local access to information on the clause type. As (6.16c) shows, the only local information of stranded prepositions is that they lack a local complement

[8] Both these pied-piped structures will have to be licensed by the *wh-pied-piping* construction.

and are therefore SLASH-ed. Consequently, a lexical stranded-preposition construction cannot encode the restriction that it is in informal relative clauses that it should be used. Instead I argue that in addition to the *formal-finite-wh-pp-filler-relative-cause* construction, British English also has an *informal-finite- wh-*$P_{stranded}$*-filler-relative-cause* construction with the following constraint:

(6.17) $\textit{fin-wh-}P_{\textit{stranded}}\textit{-fill-rel-cl} \Rightarrow [\text{REGSTR } \textit{informal}]$

(6.18) $\textit{fin-wh-}P_{\textit{stranded}}\textit{-fill-rel-cl} \Rightarrow [\text{HEAD} \;\; \textit{prep}] \rightarrow \begin{bmatrix} \text{SLASH} & \{\boxed{1}\} \\ \text{ARG-STR} & <\boxed{1}> \end{bmatrix}$

The constraint in (6.17) states that finite-*wh*-relative clauses with a stranded preposition are considered informal. In addition to this, (6.18) ensures that it is a preposition that enters the SLASH value into the clause, i.e. that is stranded. All other properties of this construction are inherited from other types: the SLASH Amalgamation will ensure that the SLASH feature of the preposition will be inherited by all mother nodes until the appropriate filler is identified. Then the head-filler construction will ensure that the SLASH feature is cancelled off.

For British English I thus propose two relative clauses dealing with preposition placement: the *informal-finite-wh-*$P_{stranded}$ *-filler-relative-clause* construction and the *formal-finite-wh-*P_{piped} *-filler-relative-clause* construction (a.k.a. Wilcock's *finite-wh-pp-filler-relative-clause* type, which I relabelled simply so as to have similar terms for the two constructions in question).

All of the above modifications (the introduction of two preposition placement constructions and two *that*-relative clauses) of Sag's 1997 relative clause construction network have been incorporated in Figure 6.12 to give the full set of constructions required for British English. The figure exemplifies how the introduction of only four constructions can add great complexity to a

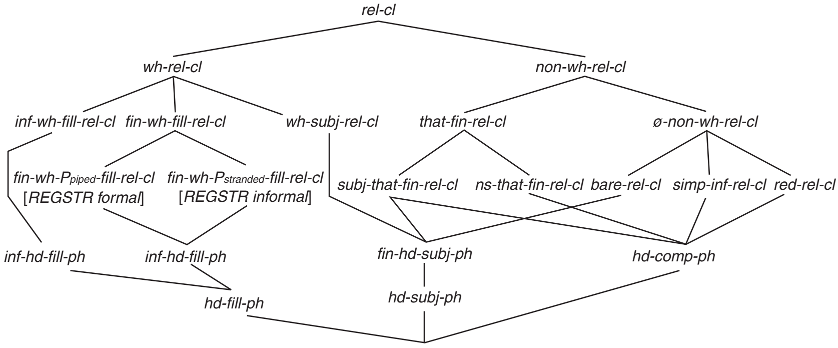

Figure 6.12 *Revised relative clause construction network for British English*

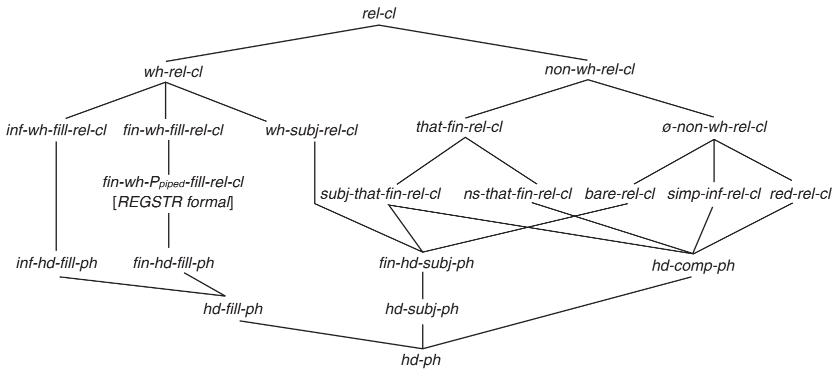

Figure 6.13 *Revised relative clause construction network for Kenyan English*

network. Irrespective of the validity of the proposed analyses, it needs to be emphasized, however, that an explicit formalization of all the underlying assumptions at least allows the above claims to be falsifiable.

In contrast to British English, the corpus results on the effect of formality in relative clauses in Kenyan English indicate that the latter variety does not have the *informal-finite-wh-$P_{stranded}$-filler-relative-cause* construction. As the Goldvarb results proved, even in informal contexts pied-piping was favoured in the ICE-EA corpus. I therefore assume that stranded *wh*-relative clauses in Kenyan English have to be constructed out of several independent constructions (namely the SLASH- and *finite-wh-filler-relative-clause* construction), while the *formal-finite-wh-P_{piped}-filler-relative-clause* construction is stored just as in British English (cf. (6.15)). Thus British English has an informal alternative to the *finite-wh-P_{piped}-filler-relative-clause* construction which will be chosen in informal contexts. In Kenyan English, on the other hand, the constructed stranded alternative structure is not marked as informal. This means that in Kenyan English in informal contexts there is not such a strong competitor to the stored *finite-wh-P_{piped}-filler-relative-clause* construction as in British English. As a result, the pied-piping preference in less formal contexts is only decreased in Kenyan English (and not reversed to stranding as in British English).

The Kenyan English relative clause construction network is thus slightly less complex than the British English one (see Figure 6.13). As the reader will have noticed, *it*-cleft-relatives are missing in the above discussion. This has to do with the fact that, as argued in section 3.1.1, these combine properties of topicalized and relative clauses.

(6.19)	a.	It was	[**John**]$_{NP_{top}}$	[**who** I talked **to**]$_{RC}$
	b.	It was	[**John**]$_{NP_{top}}$	[**to whom** I talked]$_{RC}$
	c.	It was	[**to John**]$_{PP_{top}}$	[**that** I talked]$_{RC}$

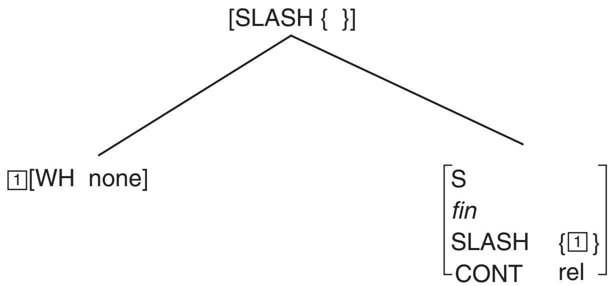

Figure 6.14 *Partial description of* it-*clefts*

As (6.19) shows, in all three *it*-clefts the topicalized element (*top*) functions as a kind of filler for the gap in the relative clause despite the fact that the relative clause already contains a *wh*-filler. There is, however, also a significant difference between (6.19a) and (6.19c), on the one hand, and (6.19b), on the other: while in the former the syntactic category of the preposed element is identical with that of the gap (i.e. it is an NP and PP, respectively) in (6.19b) the topicalized element is an NP but the gap is a PP. Clearly much further empirical research is needed on the distribution of the three alternatives in (6.19). Tentatively, however, I claim that (6.19b) can be captured by assuming an *it*-cleft construction which consists of an ordinary relative clause which modifies the topicalized antecedent. In contrast to this, I suggest the construction in Figure 6.14 for the examples (6.19a) and (6.19c). As can be seen in that figure, *it*-clefts like (6.19a) and (6.19c) are argued to consist of a topicalized filler $\boxed{1}$ which is not a *wh*-word (i.e. [WH none]). Figure 6.14 then combines properties of Ginzburg and Sag's *head-filler-phrase* (2000: 364) and *topicalised clause* (2000: 379) constructions: the non-*wh*-filler cancels off the co-indexed SLASH feature. As usual, this co-indexation ensures that the filler will have the same syntactic and semantic properties as the SLASH-ed gap in the relative clause (thus modelling the behaviour of (6.19a) and (6.19c)). Note furthermore that the fact that the relative clause still contains a SLASH feature is an idiosyncratic property of cleft-relatives that will override the default SLASH cancellation mechanism of ordinary relative clauses. Admittedly, this analysis is somewhat ad hoc and might require revision. Nevertheless, while ordinary topicalized clauses could not be investigated in the multivariate statistical analyses (cf. section 4.3), cleft-relatives at least allow a limited view on preposition placement in this special type of topicalization.

At least for British English the data, indicated, for example, that these are also subject to formality effects (cf. section 4.3.2), which might also lead to the postulation of specific formal pied-piped and informal stranded cleft-constructions in this variety. Kenyan English did not seem to require such constructions, though the low token size means that any generalization based on these data should be treated with care (cf. section 4.3.2).

6.2.3 Categorically stranding contexts

The preceding chapters have covered all constructions that are necessary for an HPSG-based Construction Grammar account of variable preposition placement in British and Kenyan English. Besides this, the discussion of *that*- and Ø-relative clauses has already illustrated how these categorically stranding environments can be explained in such an approach. This leaves the following categorical clause phenomena to be explained:

(6.20) a. [**Pied-piping**]$_i$ has been talked **about**$_i$ enough. [passive]
b. He talked about **the same stuff**$_i$ as [I talked **about**$_i$]. [comparative]
c. **His ideas**$_i$ were easy [to find fault **with**$_i$]. [hollow]
d. **What**$_i$ he is talking **about**$_i$ is called stranding. [free relative]

Passives (6.20a), comparatives (6.20b), hollow clauses (6.20c) and free relatives (6.20d) all lead to categorical stranding. Throughout this study passives have been classified together with the other three constructions in (6.20) as a clause type with a categorical effect. Yet, while this had no consequences for the analysis so far, this analysis is somewhat of a simplification: in fact, it would have been more correct to say that clauses in the passive voice obligatorily lead to preposition-stranding. *Wh*-interrogative or *wh*-relative clauses, for example, can also be constructed in the passive voice and then require obligatory stranding (cf. (6.21a) and (6.22a) vs (6.21b) and (6.22b])):

(6.21) a. [**What**]$_i$ has been talked **about**$_i$? [interrogative]
b. *[**About what**]$_i$ has been **talked**$_i$? [interrogative]
(6.22) a. the structure [[**which**]$_i$ has been talked **about**$_i$ before]. [*wh*-relative]
b. *the structure [[**about which**]$_i$ has been **talked**$_i$ before]. [*wh*-relative]

The passive construction is thus an abstract construction that affects various clause types. Therefore instead of assuming specific *wh*-interrogative-passive or *wh*-relative-passive constructions, complete-inheritance models treat passivization as a lexical rule that applies before a verb gets instantiated in a particular clause type construction (see Müller 2006: 185–8).

In HPSG passivization is analysed as 'an operation on grammatical relations, one that "demotes" subject arguments … and, in many instances, additionally "promotes" more oblique syntactic dependents' (Pollard and Sag 1994: 119). As Tseng points out, prepositional passives such as (6.20a) share many features with the ordinary passive in English, 'which already provides a mechanism for: promoting a non-subject NP to subject position, demoting the subject NP to an optional *by*-phrase, and ensuring the appropriate morphological effects' (2006: 180). He furthermore emphasizes, however, that a crucial difference between prepositional passives and ordinary passives is the fact that the former also allow 'passivizing out of adjuncts' (Tseng

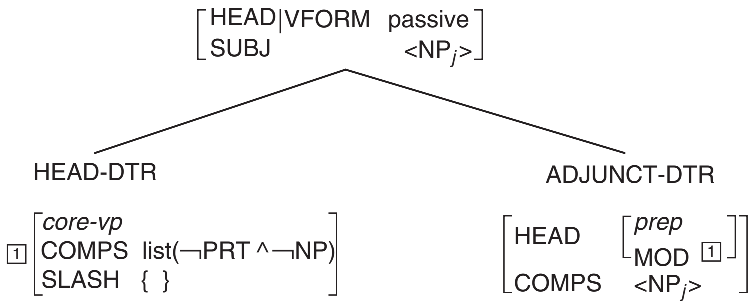

Figure 6.15 *'Constructional rule for adjunct prepositional passives' (Tseng 2006: 182)*

2006: 177; cf. the locative adjunct PP in *This chair has been sat in*). For such cases, Tseng (2006: 182) proposes the construction-based account shown in Figure 6.15. In HPSG adjuncts are considered to select for the element they modify (see Pollard and Sag 1994: 384). In Figure 6.15 this is indicated by the co-index $\boxed{1}$, which identifies the VP HEAD-DAUGHTER as the element which is modified by the ADJUNCT-DAUGHTER. The adjunct is a PP (since its HEAD is of the type *preposition*) and has an unsaturated COMPS-list (i.e. still lacks a complement specified as NP$_j$). As the mother node shows, Tseng's constructional rule guarantees that this unexpressed prepositional complement is identified with the subject of the clause (since SUBJ ⟨NP$_j$⟩; cf. 2006: 181–2).

A further requirement on the VPs in such constructions is that any particle (PRT) or NP that the verb subcategorizes for must be saturated before the adjunct PP is attached. This is encoded by the fact that the verb's COMPS list does not contain either of these two syntactic entities (for '¬' read 'not missing anymore') when the PP adjunct is added (i.e. if such elements are lexically required they must already have been realized within the VP). This constraint is meant to cover examples such as *We were made fools of* (taken from Tseng 2006: 174) or *The situation will simply have to be put up with* (taken from Tseng 2006: 176). In these sentences the NP and PRT elements are obligatorily selected by the verb and must be realized before the stranded preposition (as Tseng shows not even realizing the object NP in a non-local position is acceptable (2006: 175–6), a fact which is encoded by an obligatorily empty SLASH feature in Figure 6.15). It is not quite clear to me, however, why Tseng includes this constraint in the adjunct prepositional passive constructions. As the examples show, it is in fact prepositions that are part of lexically stored V-N-P (*make fool of*) or V-PRT-P (*put up with*) structures that require such a restriction. For adjunct PPs an intervening object NP, for example, seems to decrease the acceptability of the sentence (cf. *Someone drove a car into this wall* vs *?This wall has been driven a car into* vs *This wall has been driven into*).

$$\begin{bmatrix} \text{PHON} & \langle \textit{relied} \rangle \\ \text{HEAD|VFORM} & \textit{passive} \\ \begin{bmatrix} \text{SUBJ} & \langle \text{NP}_{\boxed{2}} \rangle \\ \text{SPR} & \langle\ \rangle \\ \text{COMPS} & \langle \boxed{4}\ \text{PP}[\textit{on}[\text{COMPS}\ \langle \text{NP}_{\boxed{2}} \rangle], (\text{PP}\ [\textit{by}, \boxed{1}\ \text{NP}]) \rangle \end{bmatrix} & \\ \text{ARG-STR} & \langle \boxed{1}\ \text{NP}, \boxed{4}\ \text{PP}[\textit{on}, \text{NP}_{\boxed{2}}] \rangle \end{bmatrix}$$

Figure 6.16 *Constructional rule for complement prepositional passives*

Tseng concludes his paper by noting 'that a similar version of the head complement rule is needed for prepositional passives involving PP complements' (2006: 182). Figure 6.16 illustrates what such a construction could look like, using the verb *rely* as an example (cf. the active voice representation in Figure 6.2. In Figure 6.16 the PP$\boxed{4}$ headed by *on*, which is obligatorily selected by *relied* (the past participle form required for passives) in its ARGUMENT-STRUCTURE, will have to be realised as a complement of the verb (cf. the verb's COMPS list). Following Tseng, however, this PP has an unsaturated COMPS-list, i.e. the prepositional complement slot NP$\boxed{2}$ is not filled. Instead this element will be identified with the subject of the clause (which is token-identical; cf. SUBJ $\langle$NP$\boxed{2}\rangle$ in Figure 6.16). (Furthermore, the agent argument is realized as an optional *by*-phrase.) Finally, although not explicitly stated in the figure, it is to this construction that Tseng's *list*(¬PRT ∧ ¬NP) constraint must apply.

Both figures imply a fundamental difference between prepositional passives and all other preposition placement constructions: while the other structures covered so far were so-called extraction phenomena involving SLASH features, prepositional passives do not employ such non-local features. A closer look at the remaining clause types shows that these, on the other hand, are modelled by SLASH-based extraction analyses:

(6.23) a. He talked about **the same stuff**$_i$ as [I talked **about**$_i$**[SLASH {$\boxed{1}$}]**].
b. **His ideas**$_i$ were easy [to find fault **with**$_i$**[SLASH {$\boxed{1}$}]**.].
c. **What**$_i$ he is talking **about**$_i$**[SLASH {$\boxed{1}$}]** is called stranding.

The comparative clause in (6.23) is syntactically similar to finite *that*- relative clauses: the NP it modifies (*the same stuff*) functions as an argument of the main clause and pied-piping is not possible (cf. **the same stuff about as I talked* or **the same stuff as about I talked*). Consequently, the SLASH feature can be bound off (cf. (6.24)) in a way similar to *nonsubject-that-finite-relative-clause* (cf. (6.13)):

$$\begin{bmatrix} \text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD} & \textit{adjective} \\ \text{SUBJ} & \langle \text{NP}_{\boxed{1}} \rangle \\ \text{COMPS} & \langle \text{VP}[\textit{inf}, \text{SLASH}\langle \boxed{2}\text{NP}[\textit{acc}]:\textit{ppro}_{\boxed{1}}\rangle]\rangle \end{bmatrix} \\ \text{NONLOCAL}|\text{TO-BIND}|\text{SLASH} \langle \boxed{2} \rangle \end{bmatrix}$$

Figure 6.17 *Lexical rule for SLASH-cancellation in* tough-*constructions (adapted from Pollard and Sag 1994: 167)*

(6.24) $\textit{comparison clause} \Rightarrow \begin{bmatrix} \text{HEAD} & [\text{MOD NP}_{\boxed{1}}] \\ \text{SLASH} & \{\ \} \\ \text{HD-DTR} & [\text{SLASH } \{\text{NP}_{\boxed{1}}\}] \end{bmatrix}$

Obviously, the constraint in (6.24) does not fully describe the semantic and syntactic properties of comparison clauses (see Borsley 2004 for potential issues of comparison clauses that still need to be addressed). However, (6.24) successfully models preposition-stranding in *as*- as well as *than*-clauses (cf. e.g. *more debts than he could cope with*), which is all that matters for the present study.

In contrast to this, Pollard and Sag (1994: 166–7) suggest that the SLASH feature in hollow clauses such as (6.23b), i.e. so-called *tough*-constructions, is cancelled off by a specific lexical rule of adjectives like *easy* or *tough*. Figure 6.17 states that if an adjective like *easy* has inherited a SLASH feature from its non-finite VP complement (cf. the COMPS list in the figure, i.e. *to find fault with* in (6.23b)) then two mechanisms apply: first of all, the gap *ppro*$\boxed{1}$ in question will be identified with the subject NP$\boxed{1}$ of the main clause. Secondly, the SLASH feature itself $\boxed{2}$ is bound off, i.e. cancelled from the remaining syntactic structure.

Finally, the last clauses that need an analysis are free relatives. These are similar to *wh*-relative clauses in that a *wh*-word (e.g. *what* in (6.23c)) is the filler for the SLASH-ed gap. In contrast to head-external relatives, however, it was pointed out in section 3.1 that in free relative clauses the *wh*-item functions as both antecedent and filler, as the paraphrase of (6.23c), shows: [*that which*] *he is talking about is called stranding*. Consequently the *wh*-item in these constructions also has to meet requirements imposed by the main clause, which is usually given as the reason why pied-piping is not possible with free relatives:

(6.25) a. *[[**About what**]$_i$ he is **talking**$_i$] is called stranding.
b. ***About that [[which**]$_i$ he is **talking**$_i$] is called stranding.

In (6.25a) *what* is the complement of the pied-piped preposition *about* in the free relative clause, which as a whole functions as the subject of the main clause. As (6.25b) illustrates, this is similar to a situation in which the preposition has been extracted out of the relative clause. Accordingly, both sentences in (6.25) are ungrammatical.

One might argue that the ungrammaticality of the examples in (6.25) is due to case mismatch: in the main clause the antecedent functions as the subject of the main clause and in the embedded clauses the *wh*-item is the complement of the preposition *about*. Yet such a view is not tenable since this description also fits the grammatical example in (6.23c). The only difference is that in the latter the preposition is stranded. In order to understand the mechanisms underlying free relative clauses it is helpful to take a look at free relative clauses in a language like German, which employs more inflectional morphemes to mark case. Müller (1999), for example, discusses the sentences in (6.26):

(6.26) a. Macht kaputt$_{acc}$, [*was*]$_{nom\_acc}$ euch kaputtmacht!
make broken what you broken.makes
'Destroy what destroys you!' (adapted from Müller 1999: 56)
b. *Er vertraut$_{dat}$, [wen]$_{acc}$ er kennt$_{acc}$.
he trusts who$_{acc}$ he knows
Intended: 'He trusts those he knows.' (adapted from Müller 1999: 62)

In both (6.26a) and (6.26b) there is a case mismatch between the requirements of the main verb and the verb of the free relative: the main verb *kaputtmachen* in (6.26a) selects *was* as its accusative object, while in the free relative the *wh*-word functions as the nominative subject. In (6.26b), on the other hand, *wen* has to function as the accusative object of *kennen* in the free relative and the dative object of *vertrauen* in the main clause. So in both clauses there is a mismatch between the case requirements of main and embedded clause verb, yet (6.26b) is ungrammatical while (6.26a) is perfectly OK. As Müller convincingly argues (1999: 85–90), sentences such as (6.26a) and (6.26b) show that it is not the structural case of a free relativizer that has to match in the main and the embedded clause. If that was the case, (6.26a) would be ungrammatical. Instead, the morphological case-marking of the free relativizer is what determines the grammaticality of these examples: in (6.26a) *was* is morphologically ambiguous in that it is both an accusative and nominative form. Due to this, *was* can meet the case requirements of both matrix and embedded verb. In contrast to this, *wen* is only accusative and thus fails to match the dative case requirement of the main verb.

Müller (1999) offers a detailed HPSG-based description of free relative clauses in German. Since this analysis can easily be applied to English free relatives, I will adopt his main assumptions, namely that clause-internally free relatives work like ordinary relative clauses. As a whole, however, this special type of relative clause functions as a nominal element in a matrix clause (via a unary projection schema which projects a relative clause into an NP). One further requirement imposed on free relative constructions then is that the list of morphological case features of the free relativizers contains information that is compatible with the main and the embedded clause.[9] The final version of the morphological case requirement then states that if the free relative clause filler is a PP, the main clause must also select for a PP (for details, Müller 1999: 91–7):

(6.27) Er schläft [mit wem] er schlafen will
He sleeps with whom he sleep wants
'He sleeps with whoever he wants to sleep with'

In (6.27) the free relativizer *wem* is embedded in a PP headed by *mit*. Since *schlafen*, which selects for a PP headed by *mit*, is the main verb in main and embedded clause the structure in (6.27) is grammatical.

Applying this analysis to English free relatives it becomes clear why (6.25a) is ungrammatical: the pied-piped PP *about what* is an acceptable filler for the free relative clause gap, but the passive verb *be called* requires an NP in subject position. Yet this analysis also predicts that pied-piping should not be completely ruled out in free relative clauses, and this is in fact what the results of the ICE corpus studies seemed to imply (cf. section 4.2):

(6.28) a. 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>
b. They are professing it **in whichever way they want to** <ICE-EA:S1A006K>

The important thing to note about the British and Kenyan examples in (6.28) is that in both cases 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 in (6.28a) and (6.28b) being grammatical.

Further experimental evidence is clearly needed to describe preposition placement in English free relatives more adequately, but it should also be kept in mind that in the multivariate corpus studies free relatives were shown to pattern with main questions and embedded interrogatives. This finding will become particularly important for the usage-based account of preposition placement in section 6.3.

Before turning to the usage-based Construction Grammar analysis, however, there is one final piece of empirical evidence that also needs to be accounted for in a complete-inheritance model. While all processing-related factors (e.g. the preference of stranding with prepositional verbs, the ungrammaticality of stranding with manner adjunct PPs or the pied-piping preference of NP-contained PPs) require no explanation within

[9] For a discussion of exceptions to this rule, see Müller 1999: 60–2.

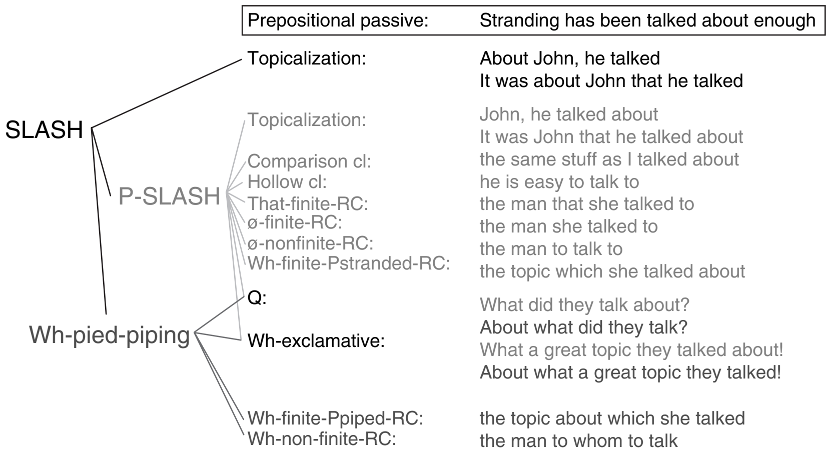

Figure 6.18 *Complete–inheritance HPSG-based Construction Grammar account of preposition placement in British English*

complete-inheritance models, the idiosyncratic properties of particular prepositions such as, *like* must be incorporated into the grammar (an observation I owe to Gert Webelhuth, p.c.). The lexicon entry for obligatorily stranded prepositions such as *like* must thus include information that in clauses with variable preposition placement *like* must be SLASH-ed (i.e. [SLASH {$\boxed{1}$}]). Prepositions with obligatory pied-piping such as *during*, on the other hand, must be obligatorily associated with the *wh-pied-piping* construction (see Figure 6.8).

The complete-inheritance HPSG based Construction Grammar account of preposition placement presented in sections 6.1–6.3 has given rise to a complex construction network which was seen to be influenced by a great number of schematic constructions. Focusing just on the two general preposition placement constructions SLASH and *wh-pied-piping* and their effect on the lowest level of constructions assumed in such an approach, Figure 6.18 summarizes the results for British English (using less technical terms for the individual constructions to make the graph more accessible to readers only interested in a usage-based description). Passives are not considered to involve extraction of arguments in HPSG, and consequently prepositional passives required an independent construction (in fact, two constructions, one for prepositional verbs and one for PP adjuncts; cf. above). Furthermore, in Figure 6.18, I have distinguished between a general SLASH construction which is involved in all types of extraction and a special P-SLASH construction (which I take to have the form P [SLASH {$\boxed{1}$}]). As pointed out earlier, for a complete-inheritance model of British English it is obviously not necessary to postulate the existence of an independent P-SLASH construction, since all its properties can be derived from the more general SLASH construction. In Figure 6.18 this construction is thus merely used for expository purposes (though I will return to this point in the usage-based section below).

The next point to note about Figure 6.18 is that pied-piping in topicalized clauses (including *it*-clefts) is treated as an extraction phenomenon (i.e. SLASH-construction), but not one that is covered by the *wh-pied-piping* construction. Furthermore, the *wh-pied-piping* construction itself is a special extraction phenomenon and thus linked to the general SLASH-construction.

In addition to this, as it turned out, the majority of preposition placement phenomena can be handled by the interaction of other, independent constructions in complete-inheritance approaches. Nevertheless three constructions with idiosyncratic preposition-stranding or pied-piping properties had to be postulated: a *non-finite-wh-relative-clause-construction*, as well as a formal pied-piped and an informal stranded-preposition construction for finite *wh*-relative clauses (i.e. *wh-finite-*P_{piped}*-relative-clause* and *wh-finite-*$P_{stranded}$*-relative-clause*, with the latter only postulated for British English and not Kenyan English). Furthermore, it was suggested that in British English an informal stranded and formal pied-piped cleft construction might also be indicated (cf. below).

The most interesting result of Figure 6.18 is probably that even the top-down approach of complete-inheritance HPSG-based Construction Grammar does not assume a single superordinate preposition placement construction. In other words, there is no maximally abstract construction which unifies all the phenomena in Figure 6.18. Working on particle placement in English (*turn on the TV* vs *turn the TV on*), Cappelle (2006) argues that the two alternative particle placement options (PRT NP vs NP PRT) are actually 'allostructions', i.e. 'variant structural realizations of a construction that is partially underspecified' (Capelle 2006: 18). The above discussion has shown, however, that no such abstract underspecified construction can be given for preposition placement in English, since the phenomena covered by this term are too disparate and consequently elude a simple generalization (apart from the fact that they all involve prepositions). Thus, the dependent variable PREPOSITION PLACEMENT employed in the present study must be considered a theoretical linguistic construct that corresponds to no single construction in a speaker's mental construction network.

6.3 Preposition placement: The enriched usage-based construction network

The previous discussion has shown that from a complete-inheritance Construction Grammar perspective the networks for preposition placement in British and Kenyan English exhibit almost the same constructions. It was only the *informal stranded finite wh-relative clause* construction[10] that was argued to be missing from the Kenyan construction network (and possibly the two cleft constructions, but cf. below). Once a usage-based perspective is adopted, however, it becomes apparent that there are significant differences between the two varieties.

Before turning to the discussion of the usage-based analysis of preposition placement in British and Kenyan English, it needs to be remembered that the driving forces behind the entrenchment of a construction in such approaches are (type- and token-) frequency and preemption. High token-frequency means that a particular substantive lexicalization of a construction occurs much more often than other lexicalizations. As a result the frequent lexicalization will be stored as a substantive construction. Productive abstract schemata, on the other hand, are only acquired, i.e. deeply entrenched, if a particular construction is encountered with many different lexicalizations. An immediate hypothesis following from this is that first-language speakers should possess more substantive as well as abstract schematic constructions in their mental grammar than second-language learners, since the latter normally receive much less input of the target language than a native speaker. Furthermore, following Hawkins (2004), I take it that processing factors also play an important role in the formation of abstract schemata. If the same content can be expressed by two competing structures and one of these is easier to process than the other, then the simpler structure will be preferred in performance. Consequently, it will be used more often with a greater range of lexicalizations, which increases its type-frequency and ultimately leads to it being more cognitively entrenched than its alternative (cf. Hawkins 2004: 6).

Furthermore, competition between structures also entails that preemption will play an important role: if on a particular occasion one construction is used instead of its alternative, then the hearer will assume that this choice reflects a functional difference between the two structures. Ultimately, this will lead to the functional differentiation of the two alternatives (i.e. the minimization of constructional synonymy). Finally, I adopt Hawkins's Performance-Grammar Correspondence Hypothesis, i.e. I take it that '[g]rammars have conventionalized syntactic structures in proportion to the degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments' (Hawkins 2004: 3). The statistically significant results from the ICE corpora studies (chapter 4) and the experiments (chapter 5) will therefore be used as evidence for the entrenchment of substantive and schematic constructions.

10 The technical term for this construction is *wh-finite-*$P_{stranded}$*-relative-clause*; for the sake of those readers who skipped the preceding sections I will use terms in this section that are more intuitively accessible but still allow the identification of the respective construction in the inheritance networks above.

The first major difference between usage-based and complete-inheritance approaches concerns the status of the stranded-preposition construction P-SLASH in Figure 6.18: as I have repeatedly pointed out, from a complete-inheritance point of view this node does not qualify as an independent construction since its properties can be derived from the more general SLASH extraction. However, since usage-based approaches adopt a bottom-up, input-driven view on schematization, it will only be via positive input that learners will acquire preposition-stranding. Since German does not license stranded prepositions, first-language learners of this language do not possess this construction. When English learners, however, are exposed to stranded prepositions they will add a schematic P-SLASH construction to their construction network. The general SLASH construction will then only arise as an abstraction of a set of more specific SLASH constructions (such as P-SLASH, NP-SLASH etc.). (In fact, from a usage-based perspective it is not even certain that such completely abstract constructions are generalized at all, but this issue is beyond the scope of the present study.)

The first phenomenon that can be considered an interplay of preemption, processing factors and type frequency was the effect of the factor group CLAUSE TYPE in the corpus study (cf. section 4.3.2). Earlier it was argued that the interaction effect of the level of formality and preposition placement in relative clauses warranted the postulation of independent constructions (a *formal-pied-piped-finite-wh-relative-clause* construction for both varieties and an additional *informal-stranded-finite-wh-relative clause* construction for British English). In addition to this, the British English data proved that cleft-relatives also exhibited a formality effect, which potentially required two further constructions for this variety. On top of that, however, the statistical analyses of both sets of data revealed that stranding was furthermore strongly favoured in questions (both main and embedded ones) and free relative clauses. While this does not lead to the introduction of an extra set of constructions in a complete-inheritance model, from a usage-based point of view it can be taken as evidence that a *stranded wh-question* and a *stranded free relative clause* construction are also part of the construction network of both varieties.

Figure 6.19 shows the consequences of this for the British English construction network. Boxes indicate constructions already postulated due to their idiosyncratic properties, while circles mark constructions which are claimed to be stored from a usage-based perspective. Two aspects of this figure need to be emphasized: first of all, the circled constructions have no additional semantic or syntactic properties and thus can easily be generated by the complete-inheritance model presented earlier (which means that in the following the internal syntax of these constructions will not have to be discussed). Secondly, as the links in the figure indicate, the existence of a stored *stranded-question* (Q_{strand}) construction, for example, does not entail that pied-piping in questions is impossible. A pied-piped question can still

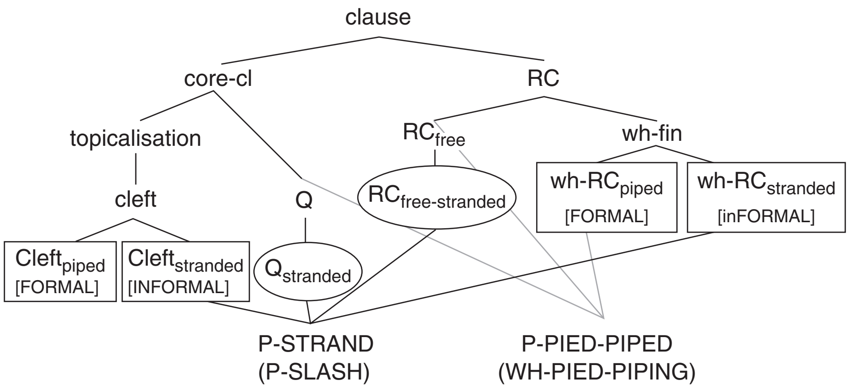

Figure 6.19 *Usage-based enriched construction network for British English*

be generated by the combination of the P-*Pied-Piped* construction and the *question* construction.[11] The intuition behind this is that constructions that are stored independently are produced more easily and can also be processed faster. Nevertheless, the alternative structure can still be constructed online should need be (see below for such a scenario involving adjunct PPs).

The corresponding construction network for Kenyan English is given in Figure 6.20. As this figure shows, Kenyan English has a slightly less complex construction network, though the absence of the two types of cleft constructions which Figure 6.19 provides for British English might be due to the low token-size of this phenomenon in Kenyan English (the ICE-EA only contained five such tokens; cf. Table 4.16). In fact, once the ICE-EA data is subjected to an independent Goldvarb analysis, it turns out that in the Kenyan data cleft-relatives behave like questions and free relatives in that they appear to favour stranding. In Figure 6.20 this finding is tentatively incorporated by including a stranded cleft-relative construction in the Kenyan network. Nevertheless, this hypothesis definitely requires further empirical corroboration (due to the low token-number on which it is based). Besides, it was argued above that Kenyan English only has a *formal-pied-piped-finite-wh-relative-clause* construction. As Figure 6.20 shows, it is still possible to construct stranded finite *wh*-relative clauses online, but since the pied-piped alternative is stored it will become available more quickly even in less formal contexts.

11 Again I am simplifying here since it is actually a combination of *nonsubject-wh-interrogative-clause*, *wh-pied-piping* and *head-filler-phrase* construction; cf. Figure 6.9. For the sake of readability, however, I will use more accessible non-technical terms in the following that still allows identification of the technically more precise constructions in the preceding sections.

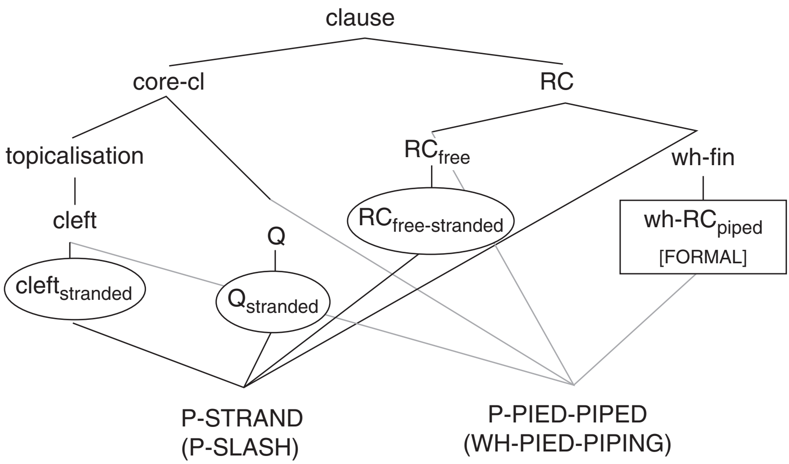

Figure 6.20 *Usage-based enriched construction network for Kenyan English*

Note how the entrenchment of a *stranded-question* construction in both varieties can be considered the result of processing factors: as pointed out in section 3.1.3, the function of a *wh*-word in questions is to introduce new information and signal interrogation. Consequently it should appear clause-initially, something which can be guaranteed by stranding the preposition. In relatives, on the other hand, a pied-piped preposition adds information as to the syntactic and semantic function of the *wh*-word and facilitates the interpretation of the entire relative clause. Furthermore, pied-piping has the advantage of obeying Hawkins's (2004) 'Avoid Competing Subcategorizors' and 'Valency Completeness' principles (i.e. it avoids garden-path effects and allows integration of a filler upon encountering the main verb of a clause; see section 3.5). Since processing effort is greatest in relative clauses (also see section 3.5), it thus not surprising that pied-piping should be more frequent in these clauses, leading to the entrenchment of this construction. These effects thus explain why in Kenyan English a *stranded-question* and a *pied-piped-wh-relative-clause* construction are stored.

It was, furthermore, argued above that first-language speakers receive more input. This higher type-frequency of preposition-stranding and pied-piping together with the effect of preemption then explains why British English has an extra *informal-stranded-finite-wh-relative-clause* construction: British speakers encounter more stranded prepositions in general, and accordingly also in online-constructed finite *wh*-relative clauses, which leads to the structure *wh*- + P being available for storage. Due to the influence of preemption, however, storage is facilitated greatly if a structure exhibits a functional idiosyncrasy that sets it apart from its structurally similar alternative. Since P + *wh*-relative clauses are already marked as formal, the straightforward idiosyncratic property that stranded *wh*-relative clauses are associated with is thus informality.[12]

Hawkins has claimed repeatedly (1999, 2004) that all things being equal pied-piping should be preferred over stranding since it satisfies both the 'Avoid Competing Subcategorizors' and the 'Valency Completeness' principles. As pointed out in section 3.1.4, this hypothesis is certainly supported by the cross-linguistic preference of pied-piping. Yet, once a language has both preposition placement options it is not the case that pied-piping will always be the simpler choice. The corpus data for British and Kenyan English have, for example, illustrated that there are situations in which stranding can be considered to be preferred over pied-piping from a processing perspective (e.g. interrogatives favouring stranding since this leaves the question word to introduce the clause). Furthermore, Hawkins himself emphasizes that stranding will be preferred for prepositions which 'are highly dependent on verbs for their interpretation and processing' (Hawkins 1999: 260, fn. 15). In contrast to this, more adjunct-like PPs are expected to function as a single unit and therefore favour pied-piping.

These claims received strong support from the empirical data: in the corpus studies of both varieties,

(1) the PP types which are most closely associated with a particular verb, i.e. prepositional verbs and V-X-P idioms, were strongly associated with those clause types that induce categorical stranding (passives, non-*wh*-relatives, hollow and comparison clauses; see section 4.2.3);
(2) for all variable clause types, prepositional verbs and V-X-P idioms were identified as favouring stranding most (see Table 4.21).

In addition to this, the experimental results indicated that for both varieties the stranding preference with prepositional verbs increases with increasing complexity of the filler–gap domain in both varieties (see section 5.2).

In contrast to this, more adjunct-like PPs, such as temporal or locational adjunct PPs, exhibited the expected pied-piping preference:

(1) in the ICE corpora they were significantly less frequent than expected by chance in the categorical stranding clauses (see section 4.2.3), and
(2) in the variable data they always disfavoured stranding the most (cf. e.g. section 4.3.2).

Finally, this was also corroborated by the experimental data, in which structures with stranded adjunct PPs consistently received judgements lower than those with stranded prepositional verbs (see sections 5.1 and 5.3).

[12] The informality of stranded prepositions then also seems to have spread to Ø-relative clauses, which an HCFA proved to be significantly associated with informal text types in British English (see Table 4.30). Since in Kenyan English no informal *wh*- + P model exists it is not surprising that there Ø-relative clauses occur more often in formal contexts instead (also see Table 4.30).

From a usage-based perspective, the orthogonal effect of prepositional verbs and prototypical adjunct PPs with respect to preposition placement can be explained fairly easily: prepositional verbs like *rely on* often have a non-compositional meaning and must be stored in the mental lexicon. Once learners have acquired such verb–preposition structures they require positive input in order to know that these lexical items can actually be separated syntactically by an intervening adverbial (e.g. *Rome also* ***[relied] [more and more] [on provincials]*** <ICE-GB:W2A-001 #87:1>) or that the preposition can even be pied-piped away from the verb to the front of a clause (e.g. Lord Whitelaw the man **[on whom] Mrs Thatcher relied** for so long <ICE-GB:S2B-009 #59:1:E>; see Quirk *et al.* 1985: 1163). In prototypical adjunct PPs such as the locational adjunct *in London*, on the other hand, the preposition has closer ties with its complement and the whole PP normally has the function of a modifier. Consequently for these structures the learner requires positive input that such PP complexes can be broken up, e.g. *Which room did he die in*?. One piece of evidence for the role of input frequency is the fact 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: in ICE-EA there were more adjunct PPs that exhibited categorical pied-piping than in the ICE-GB data (e.g. with instrument or cause / reason / result adjunct PPs; see section 4.3.2). Moreover, while stranded locational and temporal adjuncts PPs received acceptable ratings in relative (5.1.1) and interrogative clauses (5.3.1) by the British speakers, the Kenyan speakers strongly preferred pied-piping in these contexts. It was therefore concluded that preposition-stranding is less productive and limited to fewer contexts in Kenyan English.

As the corpus study showed, however, the degree of closeness of verb–preposition structures must be conceived of as a cline and not a simple complement–adjunct dichotomy. In between prepositional verbs and prototypical adjuncts lie for example more complement-like PPs such as optional complement PPs (e.g. *she talked about/to/with him*), which favour stranding, and obligatory complements (*they lived alone/in New York/happily ever after*), which seem to show no preference for either stranding or pied-piping. The effect in each of these cases, however, is less pronounced than that of the two prototypical endpoints of the scale.

Furthermore, there is a set of adjunct PPs, namely respect, manner, frequency and degree, which only appear pied-piped in both corpora (see section 4.3.1.4). It was argued that this was due to the fact that these PPs do not add simple thematic participants to predicates but must be interpreted as complex semantic functions (e.g. manner adjuncts, comparing events to other similar events). Consequently, preposition-stranding with these PPs was considered to be eschewed since the resulting structures in these cases seriously impeded the semantic interpretation of the PP. This hypothesis was supported by the experimental data, which identified preposition-stranding with manner and frequency adjuncts as a soft, i.e. semantic, constraint violation (sections 5.1 and 5.3). Since speakers thus only encounter these PP adjuncts with pied-piped prepositions, it seems reasonable to assume that a usage-based construction network will also contain specific pied-piped manner, respect, degree and frequency constructions (i.e. versions of the *wh-pied-piping* construction in Figure 6.8 whose semantic content value is specified as a modifier of manner, frequency etc.).

The different effects of the various PP types obviously also have repercussions for the construction network of British and Kenyan English. As Gries (2003: 157–84) shows, it is possible to interpret significant effects of multivariate studies as probabilistic cognitive constraints. In other words, factors that turn out to statistically influence the choice between two competing constructions (e.g. the effect of PP types on preposition-stranding versus pied-piping) can be seen as correlates of cue-strength in cognitive connectionist/interaction activation models (see e.g. Bates and MacWhinney 1989; Rumelhart, McClelland and the PDP research group 1986). Now usage-based Construction Grammar approaches straightforwardly allow for the incorporation of such probabilistic mental knowledge since entrenchment is a gradual phenomenon: the higher the input frequency of a particular construction, the stronger it is going to be entrenched in the neural network.

Based on the results from the Goldvarb analysis of the corpus data (Table 4.21), Figure 6.21 illustrates how such probabilistic preferences can be encoded in the PP types construction network. The factor weights of the Goldvarb analysis are represented as follows: weak effects favouring either variant with weights above 0.5 and below 0.6 (log odds $0 < x < 0.405$) are given as single lines and for each 0.1 increase in the factor weights an extra line is added (thus weights above 0.6 and below 0.7 (log odds $0.405 < x < 0.847$) are indicated by two lines, while three lines represent weights above 0.7 and below 0.8 (log odds $0.847 < x < 1.386$), etc.). In a similar way, moderate inhibiting effects (i.e. weights above 0.4 and below 0.5 / log odds $-0.405 < x < 0$) are also given as a single line (since favouring and inhibiting weights close to the threshold value of 0.5 only have weak effects). In contrast to this, stronger inhibiting effects are indicated by increasingly broken lines.[13]

The statistical corpus analysis showed that PP types in both varieties exhibit similar effects, so that Figure 6.21 can be seen as a partial representation of the PP construction network of both British and Kenyan English. Yet, due to the different type and degree of input, the actual L1 and L2 networks cannot, of course, be expected to be perfectly identical. As mentioned

13 Table 4.21 only gives factor weights with reference to the effect of factors on preposition-stranding. Due to the binomial nature of the dependent variable, however, the pied-piped weights are simply $1 - \text{weight}_{\text{stranded}}$. Note furthermore that the cut-off point for the above categories is admittedly arbitrary and that Figure 6.21 is obviously only intended as a schematic representation. For the precise effects of the various PP types see the exact Goldvarb weights and log odds in Table 4.21.

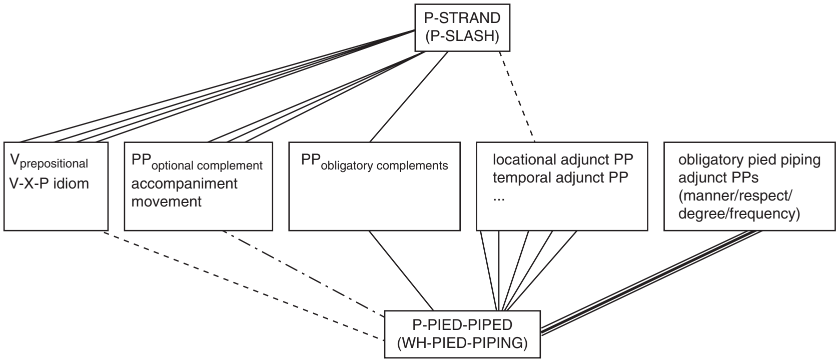

Figure 6.21 *Usage-based construction network of various PP types (British and Kenyan English)*

above, for example, the results from the categorical stranding constructions (section 4.3.2) or the experiments presented in sections 5.1 and 5.2 confirm 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. Consequently, due to more and varied input both the *P-Pied-Piped* and the *P-Strand* construction can be said to be more deeply entrenched in L1 British English. On top of that, Figure 6.21 illustrates that, from a usage-based perspective, both varieties can be said to possess stored *obligatory pied-piping (manner, respect, degree and frequency) adjunct PP* constructions. As will be remembered from section 4.4.2, however, due to more input, British English also has, for example more specific antecedent + P relative-clause structures than Kenyan English.

Another processing-based factor which surfaced in the corpus data was the type of phrase in which the PP is embedded. As the Goldvarb analysis showed, the main effect of this factor group concerned the strong pied-piping preference of NP-contained PPs (see section 4.3.2). In section 3.4 it was stressed that regardless of whether the preposition is stranded or pied-piped, NP-contained PPs always violate Hawkins's 'Valency Completeness' principle since in both cases the human processor has to look beyond the main subcategorizor to integrate the filler. Yet, due to the fact that a pied-piped preposition allows a slightly earlier integration of the filler this variant is preferred over stranding (which requires the processor to integrate the filler with a gap that is contained in a stranded preposition that itself is embedded in an NP). From a usage-based perspective it thus seems possible that this strong statistical preference for pied-piping thus leads to the entrenchment of an abstract [N [PP{SLASH}]$_{NP}$ construction.

The final factor group that affected preposition placement in Kenyan English as well as British English was the restrictiveness of a relative clause (Table 4.28). For both varieties, non-restrictive relative clauses favoured preposition-stranding (with a factor weight of 0.729 / log odds: 0.652) while restrictive relative clauses favoured pied-piping (inhibiting stranding with a factor weight of 0.422 / log odds: -0.652). This also can be interpreted as a processing effect: non-restrictive relative clauses are not necessary for the identification of the reference of the antecedent NP. Consequently, the filler–gap identification process in non-restrictive relative clauses is less complex than in restrictive relative clauses (see Hawkins 2004: 240–2). This reduced complexity in non-restrictive relative clauses then allows the use of stranding, which in itself involves more processing load than pied-piping (see section 4.2).

From a complete-inheritance-model perspective there thus exists no need to postulate an independent set of constructions to explain this effect. From a usage-based point of view it could be argued, however, that in non-restrictive relative clauses*wh*-relativizers occur more frequently in contexts which in restrictive relative clauses favour both stranding and *that*/*Ø* (which are banned from non-restrictive relative clauses). As a result, the factor non-restrictive itself might become interpreted as favouring stranding, leading to the entrenchment of an extra construction. However, the empirical support for this claim at current is rather weak, since a look at the raw frequencies reveals that 73.8 per cent (= 135/183) of all non-restrictive *wh*-relative clauses have a pied-piped preposition. Without further evidence to the contrary, it is therefore more convincing to ascribe this effect to processing factors only, and not to an independently stored construction. (Though this, of course, raises the interesting question of how subtle the statistical input information has to be in order to lead to cognitive entrenchment – something that definitely needs to be investigated in future research.)

So far the above discussion has focused on abstract schemata and their entrenchment in British and Kenyan English, but throughout the corpus studies several (partly) substantive constructions were also identified: the collocation *be like*, for example, whose preposition exerts an idiosyncratically stranding effect, was seen to be significantly more frequent in British English than in Kenyan English (see Table 4.12). In a usage-based approach the statistically higher token-frequency of this particular structure can be interpreted as a sign that this substantive construction is more deeply entrenched in British English. Furthermore, only the ICE-GB data contained an apparent exception to the rule that frequency adjunct PPs categorically pied-pipe prepositions. As it turned out, the tokens in question all contained the partly substantive, discontinuous *how…for* construction that due to its storage in the construction network (see Table 4.15) exceptionally licenses stranded prepositions with frequency PPs in British English. Finally, in section 4.4.2 it was shown that in both varieties the set of categorically pied-piping PP types (manner, degree, respect and frequency PPs) exhibit various antecedent + P structures in relative clauses that turned out to be significantly associated. Consequently, complex, partly substantive relative clause constructions can be postulated that have a filled antecedent and a clause-initial P-slot (e.g. *extent to*, which was significant for both British and Kenyan English; see Table 4.27).

As with the more schematic constructions above, the higher input-frequency of phenomena that first-language speakers are exposed to also leads to a deeper entrenchment of the more substantive constructions: both the *be like* and the *how…for* constructions are more deeply entrenched in British English. Moreover, while the statistical analysis identified eleven significantly associated antecedent + P constructions in the British English categorically pied-piping PP tokens, the Kenyan English data included only three such constructions.

Finally, the above-outlined usage-based Construction Grammar approach can also account for a doubled-preposition structure such as (6.29):

(6.29) It has resulted in late transposition and hurried implementation with little useful guidance in place **on which business can rely on**. (www.parliament.the-stationery-office.co.uk/pa/cm200405/cmselect/cmenvfru/102/4111709.htm)

Example (6.29) is from an official document of the Select Committee on Environment, Food and Rural Affairs of the British Parliament. Obviously in (6.29) it is the conflicting effect of two strongly activated constructions that leads to preposition-stranding: the very formal setting leads to strong activation of the *formal-wh-finite-*P_{Piped}*-relative-clause* construction. In addition to that, however, a prepositional verb like *rely on* (cf. Figure 6.21) strongly activates the stranded alternative as well. Thus, the doubled preposition in (6.29) can be interpreted as a blend of the following two simultaneously activated constructions (this analysis was suggested to me by Joseph Hilferty, p.c. via the WORDGRAMMAR mailing list):[14]

(6.30) a. **on** [which business can rely]
b. [which business can rely] **on**

As (6.30) shows, the two constructions differ only with respect to the position of the preposition, sharing all other syntactic material. Since both constructions are furthermore strongly activated, blending them can occur occasionally (though normally one of the two constructions will be activated more strongly, thus leading to the realization of only one of the two canonical structures).

In addition to this, it was argued that both prepositional placement variants are locally acceptable, thus combining the processing advantages of preposition pied-piping and stranding, but that the construction as a whole is not part of the constructional network of either British or Kenyan English.

[14] www.jiscmail.ac.uk/lists/WORDGRAMMAR.html, 29 June 2005.

Yet, most Kenyan English speakers will posses a construction network that is less deeply entrenched than that of L1 British English speakers (as also evidenced e.g. by lower judgement scores for all stimuli by Kenyans in the second experiment; see section 5.2.2). 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 appear more frequently in the second-language learner's data (section 4.3.1.1) and receive better scores from Kenyan English speakers (sections 5.2.2; 5.3.2).

In this chapter I have tried to give an empirically adequate Construction Grammar analysis of preposition placement in British and Kenyan English. For this, I first presented a complete-inheritance HPSG-based Construction Grammar analysis that ensured that all postulated constructions interacted with all other units of a constructional network without causing any unexpected, hidden computational problems. Then I illustrated how such a complete-inheritance network can be enriched by usage-based information. I am aware that such an approach has, of course, a top-down flavour to it that might be deemed undesirable from a bottom-up usage-based perspective which puts emphasis on how constructions are gradually acquired. However, in order to provide such an account of preposition placement acquisition, longitudinal empirical studies for both varieties are required. Studies on adult language such as the present one, on the other hand, should be seen as outlining the system which is the result of these acquisition processes.

7 Conclusion: The verdict

> This is something up with which I will not put. (attributed to Sir Winston Churchill; Pullum and Huddleston 2002: 629)

The above sentence is a much-quoted joke attributed to Sir Winston Churchill, who is supposed to have commented on a particularly stilted evasion of preposition-stranding with the words 'This is something up with which I will not put' (Pullum and Huddleston 2002a: 629). The sentence is intentionally ungrammatical since not only the preposition *with* has been pied-piped to the front of the clause, but also the particle *up*. Interestingly, though the sentence is clearly ungrammatical, it provides another piece of evidence for Construction Grammar approaches.

Searching for sentences similar to the quote on British internet sites on Google,[1] for example, gives sixteen examples of the structure, nine of which have been adjusted in the following ways (emphasis, i.e. bold font, italics and brackets, added):

(7.1) a. Inconsistency is
[*the sort of infelicity*] **up with which** [*we*] **will not put**.
(users.ox.ac.uk/~jrlucas/Godel/satan.html)
b. In Nottingham this is [*a situation*] **up with which** [*we*] **will not put**!
(voxx.demon.co.uk/eccent/ eccentd.php?filename=00000088.txt)
c. I am sure that, if there is a great deal of disruption in Victoria street, the formidable Lady Porter will be banging on somebody's door in order to let the Minister know that that is
[*the sort of thing*] **up with which** [*she*] **will not put**.
(www.publications.parliament.uk/pa/cm198990/cmhansrd/1990–05–22/Debate-10.html)
d. But, faced with a choice between Ruweished and the 'chaos' and 'insecurity' of Iraq, the Palestinians have finally found
[*a refugee camp*] **up with which** [*they*] **will not put**.
(www.telegraph.co.uk/.../news/2003/06/01/wsteyn01.xml&sSheet=/news/2003/06/01/ixnewstop.html)

[1] The search was conducted by looking for the string 'up with which * put|puts|putting' on British '.uk' Internet sites on 27 March 2005.

e. I hope that the Minister will tell the Chancellor that this is **something up with which [*she*] will not put** (www.publications.parliament.uk/pa/cm200304/cmhansrd/cm040302/halltext/40302h01.htm)
f. Baroness Blatch: My Lords, I have to agree with the right reverend Prelate. This is a scourge involving [*the exploitation of children*] **up with which [*we*] [*should*] not put**. (www.parliament.the-stationery-office.co.uk/pa/ld199596/ldhansrd/vo951207/text/51207–01.htm)
g. one should have the honesty to say simply that some things are [*metaphysical nonsense*] **up with which [*one*] will not put.** (www.wittgenstein.internet-today.co.uk/strathern.html)
h. This is [*the sort of wishy-washy thinking*] **up with which [*purists*] will not put**. (archive.thisisyork.co.uk/2005/1/19/231676.html)
i. to the preservation of the English language, and for drawing your readers' attention to [*our shoddy apostrophe*], **up with which [*he*] will not put**. (archive.thisisworcestershire.co.uk/2001/11/17/298048.html)

The first thing to note is that among the most variable slots of the sentence is the antecedent position: only (7.1e) has *something* as in the original quote. In addition to that, the subject position is also filled differently: *we* in (7.1a,b,f), *she* in (7.1c,e), *he* in (7.1i), *they* (7.1d), *one* (7.1g), and *purists* (7.1h). Furthermore, in (7.1f) the modal *should* is used instead of *will*. In contrast to this, the elements *up with which … not put* remain constant.

Obviously, I am not claiming that particle pied-piping as in the Churchill quote is becoming grammatical. Examples such as (7.1) are very limited in frequency and are also only used by a particular (educated) stratum of society (e.g. (7.1c,e,f) are from the official records of the British Parliament). What the examples show, however, is how a fully substantive idiom such as this quote can be progressively schematized by speakers. For while the sentence is ungrammatical, speakers can still identify the individual syntactic slots of the construction filled by the various substantive items. The construction that emerges from the above examples can thus be given as in (7.2):

(7.2) [$\mathrm{N}_{\mathrm{antecedent}}$] [[up with which] [$\mathrm{NP}_{\mathrm{subj}}$] $\mathrm{V}_{\mathrm{modal}}$ [not] [put]]$_{\mathrm{RC}}$

As (7.2) shows, while some speakers have now schematized the antecedent noun, subject NP and modal verb slots, the remaining parts of the quote are retained. (Note, furthermore, that the construction in (7.2) also has a prototypical discourse function associated with it, since it is mainly used to mock someone else's opinion or behaviour.)

Construction Grammar approaches can thus capture even the most peripheral of linguistic phenomena, while at the same time they are also capable of giving descriptively and explanatorily adequate grammatical analyses of all core phenomena. In this book, I have tried to show how even the complex distribution of preposition placement in L1 British English as well as L2 Kenyan English can be analysed by such an approach. Using corpus and introspection data as corroborating evidence, I was able to show that both varieties are subject to the same processing constraints (e.g. prepositional verbs favouring stranding and adjunct PPs favouring pied-piping) and must actually share a great number of constructions (e.g. the *stranded-question* or *formal-wh-finite-P*$_{piped}$*-relative clause* construction). In addition to that, however, the corpus as well as the experimental data indicated that all these constructions are less entrenched and less productive in the second-language variety. Adjunct PPs, for example, are much less strongly associated with the stranding construction in Kenyan English than in British English. Note that while this is not per se a surprising result, the present study has actually been able to support this claim by hard statistical evidence. In contrast to this, for example, it turned out that the categorical pied-piping tendency of most of the Kenyan L1 languages (notable in the set of Bantu languages), did not appear to have any effect. Instead, as expected by usage-based Construction Grammar theories, it was input and processing factors that mainly seemed to determine the construction networks of both British and Kenyan English.

Data such as (7.1) clearly indicate that even the in-depth corpus and experimental studies presented in this book were not able to exhaustively cover all preposition placement phenomena. Furthermore, individual speakers will obviously have entrenched different constructions to different degrees. Thus the relationship between individual mental grammars and the emergent construction networks of the British and Kenyan English speech communities postulated in chapter 6 requires much further research. However, as I have tried to indicate, these construction networks can easily accommodate constructions such as (7.1), simply by treating them as more substantive and lexically stored instantiations of more general, schematic constructions. For example, (7.1) can be said to share several properties with the *formal-wh-finite-P*$_{piped}$*-relative clause* construction: it is a *head-filler-phrase*, the relative clause as a whole functions as a modifier, and the construction is normally used in formal settings. It, of course, also overrides certain parts of the more general construction (most notably by employing a non-standard particle + preposition + *wh*-filler and by having a non-compositional discourse meaning), but such defaults can easily be overruled by the lexical information of substantive constructions in Construction Grammar approaches.

For researchers working within a Construction Grammar framework the present study thus offers a construction network template for British and Kenyan English which they can extend and revise as necessary. Researchers who think that Construction Grammar is something up with which any serious linguist should not put, on the other hand, can focus on the empirical results of chapters 4 and 5 and investigate the repercussions these have for whatever theory they work with.

Appendix

The Appendices to this book can be found online at: www.cambridge.org/ hoffmann. The contents of the appendices are listed below. The Appendix numbers correspond to the respective chapter and section numbers in the book. (A.4.2.3 is thus referred to and discussed in section 4.2.3.)

A.4.2.3	HCFA Obligatory Clause Types
A.4.3.1.1	HCFA Non-standard EA
A.4.3.1.3	HCFA BE tokens
A.4.3.1.5	HCFA Obligatory PPs
A.4.3.2	Multivariate Goldvarb analysis ICE-GB vs. ICE-EA
A.4.3.2.1	Original model
A.4.3.2.2	Revised models
A.4.3.2.3	Final Goldvarb model
A.4.3.2.4	Final Rbrul model
A.4.3.2.5	Cross-validation of final model in R
A.4.3.2.6	Model including factor group displaced element
A.4.4.1	HCFA Non-finite Relative Clauses
A.4.4.2	Categorical pied-piping PPs: Covarying-collexeme analysis
A.4.4.3	Variable RC data: Goldvarb analyses
A.4.4.3.1	Initial model
A.4.4.3.2	Final recoded data set
A.4.4.3.3	Other models
A.4.4.3.4	Binomial One-level of best model
A.4.4.2.5	Final Rbrul model
A.4.4.3.6	Cross-validation of best model
A.4.4.4	HCFA Finite Relative Clauses
A.5.1	Preposition placement in simple relative clauses
A.5.1.1	SPSS results of British English speakers
A.5.1.2	SPSS results of Kenyan English speakers
A.5.2	Preposition placement in relative clauses ofvarying complexity
A.5.2.1	Sample material set
A.5.2.2	SPSS results of British English speakers
A.5.2.3	SPSS results of Kenyan English speakers
A.5.3	Preposition placement in interrogative clauses
A.5.3.1	Sample material set
A.5.3.2	SPSS results of British English speakers
A.5.3.3	SPSS results of Kenyan English speakers

References

Aarts, Bas. 2000. Corpus linguistics, Chomsky and Fuzzy Tree Fragments. In: Christian Mair and Marianne Hundt, eds. *Corpus Linguistics and Linguistic Theory*. Amsterdam and Atlanta, GA: Rodopi, 5–13.

Aarts, Bas, David Denison, Evelien Keizer and Gergana Popova, eds. 2004. *Fuzzy Grammar: A Reader*. Oxford: Oxford University Press.

Aarts, Jan. 1991. Intuition-based and observation-based grammars. In: Karin Aijmer and Bengt Altenberg, eds. *English Corpus Linguistics*. London/New York: Longman, 44–62.

Abdulaziz, Mohamed H. 1991. East Africa (Tanzania and Kenya). In: Cheshire, ed. 1991: 391–401.

Aitchison, Jean. 1991. *Language Change: Progress or Decay?* 2nd edn. Cambridge: Cambridge University Press.

Allen, Cynthia. 1980. Movement and deletion in Old English. *Linguistic Inquiry* **11**: 261–323.

Allison, Paul D. 2008. Convergence failures in logistic regression. In: SAS Institute Inc. 2008. *Proceedings of the SAS® Global Forum 2008 Conference*. Cary, NC: SAS Institute Inc. www2.sas.com/proceedings/forum2008/360–2008.pdf.

Alston, R. C. 1965. *A Bibliography of the English Language from the Invention of Printing to the Year 1800*. Vol. I. Leeds: Arnold and Son.

Appleby, Lee L. 1961. *A First Luyia Grammar*. Nairobi: The East African Literature Bureau.

Ash, John. 1760. *Grammatical Institutes: Or, An Easy Introduction to Dr. Lowth's English Grammar, Designed For The Use Of Schools*. Worcester. 4th edn. Repr. in facs. by Robin C. Alston (1967): *English Linguistics* 9. Leeds: The Scholar Press Limited.

Baayen, R. H. 2008. *Analyzing Linguistic Variation: 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*. (Linguistische Arbeiten 172.) Tübingen: Max Niemeyer, 156–77.

Ball, Catherine N. 1996. A diachronic study of relative markers in spoken and written English. *Language Variation and Change* 8: 227–58.

Bard, Ellen Gurman, Dan Robertson and Antonella Sorace. 1996. Magnitude Estimation of Linguistic Acceptability. *Language* **72**: 32–68.

Bardovi-Harlig, Kathleen. 1987. Markedness and salience in second-language acquisition. *Language Learning* **37**: 387–407.

Barlow, Michael and Suzanne Kemmer, eds. 2000. *Usage-based Models of Language.* Stanford, CA: CSLI Publications.

Barrett-Keach, Camillia. 1985. *The Syntax and Interpretation of the Relative Clause Construction in Swahili.* New York/London: Garland Publishing, Inc.

Bates, Elisabeth and Brian MacWhinney. 1989. Functionalism and the competition model. In: Elisabeth Bates and Brian MacWhinney, eds. *The Crosslinguistic Study of Sentence Processing.* Cambridge: Cambridge University Press, 3–73.

Bayley, Robert and Dennis R. Preston, eds. 1996. *Second Language Acquisition and Linguistic Variation.* Amsterdam/Philadelphia: Benjamins.

Bengtsson, Per. 1996. Prepositional relatives in English: A diachronic case study of three authors. Unpublished paper. Department of English, Göteborg University.

Bentley, Muriel and Adrian Sherwood. 1964. *English for Modern Africa.* London: George Allen and Unwin Ltd.

Bergh, Gunnar and Aimo Seppänen. 2000. Preposition stranding with Wh-relatives: A historical survey. *English Language and Linguistics* **4**: 295–316.

Biber, Douglas. 1988. *Variations across Speech and Writing.* Cambridge: Cambridge University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English.* London: Longman.

Bock, J. Kathryn. 1987. Coordinating words and syntax in speech plans. In: Andrew Ellis, ed. *Progress in the Psychology of Language.* London: Lawrence Erlbaum, 337–90.

1990. Structure in language: Creating form in talk. *American Psychologist* **45**: 1221–36.

Borsley, Robert D. 2004. An approach to English comparative correlatives. In: Stefan Müller, ed. *Proceedings of the HPSG04 Conference.* Stanford, CA CSLI Publications. csli-publications.stanford.edu.

Bortz, Jürgen. 2005. *Statistik für Human- und Sozialwissenschaftler.* 6th edn. Heidelberg: Springer Verlag.

Bortz, Jürgen, Gustav Lienert and Klaus Boehnke. 1990. *Verteilungsfreie Methoden in der Biostatistik.* Berlin: Springer.

Brauner, Siegmund and Joseph K. Bantu. 1967. *Lehrbuch des Swahili.* Leipzig: VEB Verlag Enzyklopädie Leipzig.

Broccias, Christiano. 2006. Cognitive approaches to grammar. In: Gitte Kristiansen, Michel Achard, René Dirven and Francisco J. Ruiz de Mendoza Ibáñez, eds. *Cognitive Linguistics: Current Applications and Future Perspectives.* Berlin/New York: Mouton de Gruyter, 80–115.

Brown, Keith and Jim Miller. 1991. *Syntax: A Linguistic Introduction to Sentence Structure.* 2nd edn. London: Routledge.

Bukenya, Austin, Angelina Kioko and David Njeng'ere. 2004. *Head Start Secondary English: Form 3. Teacher's Book.* Nairobi: Oxford University Press.

2005. *Head Start Secondary English: Form 4. Teacher's Book.* Nairobi: Oxford University Press.

Bukenya, Austin, Angelina Kioko, David Njeng′ere and Virginia Mutei. 2003a. *Head Start Secondary English: Form 1. Teacher's Book*. Nairobi: Oxford University Press.

Bukenya, Austin, Angelina Kioko, David Njeng′ere and Jane Njue. 2003b. *Head Start Secondary English: Form 2. Teacher's Book*. Nairobi: Oxford University Press.

Burchfield, Robert W. 1996. *The New Fowler's Modern English Usage*. Oxford: Clarendon Press.

Bury, Dirk. 2003. Phrase structure and derived heads. Ph.D. thesis, University College London. www.phon.ucl.ac.uk/home/dirk/PS derivedheads.pdf.

Bybee, Joan. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam/Philadelphia: Benjamins.

1995. Regular morphology and the lexicon. *Language and Cognitive Processes* **10**: 425–55.

2000. The phonology of the lexicon: Evidence from lexical diffusion. In Barlow and Kemmer, eds. 2000: 65–85.

2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

2006. From usage to grammar: The mind's response to repetition. *Language* **82**: 711–33.

Bybee, Joan and Paul Hopper, eds. 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam/Philadelphia: Benjamins.

Cappelle, Bert. 2006. Particle placement and the case for 'allostructions'. *Constructions* SV1–7. www.constructions-online.de/articles/specvol1/683/Cappelle_Particle_Placement.pdf.

Casenhiser, Devin and Adele E. Goldberg. 2005. Fast mapping between a phrasal form and meaning. *Developmental Science* 8: 500–8.

Chalker, Sylvia and Edmund Weiner. 1994. *The Oxford Dictionary of English Grammar*. Oxford: Oxford University Press.

Cheshire, Jenny, ed. 1991. *English around the World: Sociolinguistic Perspectives*. Cambridge: Cambridge University Press.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton and Co.

1961. Some methodological remarks on generative grammar. *Word* **17**: 219–39.

1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

1981. *Lectures on Government and Binding*. (Studies in Generative Grammar 9.) Dordrecht, Holland: Foris Publication.

1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

2000. Minimalist inquiries: The framework. In: Roger Martin, David Michaels and Juan Uriagereka, eds. 2000. *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Cambridge, MA: MIT Press, 89–155.

2001. Derivation by phase. In: Michael Kenstowicz, ed. Ken Hale: *A Life in Language*. Cambridge, MA: MIT Press, 1–52.

Comprehensive Topical English. 2006. KCPE Revision Book with Reminder Notes and Reasoned Answers. 1990–2005. Standard 5–8. Nairobi: High Flyer Services.

Cook, Vivian and Mark Newson. 1996. *Chomsky's Universal Grammar: An Introduction*. 2nd edn. Oxford: Blackwell.

Cowart, Wayne. 1997. *Experimental Syntax: Applying Objective Methods to Sentence Judgements*. Thousand Oaks, CA: Sage.

Crawley, Michael J. 2005. *Statistics: An Introduction Using R*. Chichester: John Wiley and Sons Inc.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Croft, William and D. Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Culicover, Peter W. and Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press.

Davis, William D. and Stanley Dubinsky. 2003. On extraction from NPs. *Natural Language and Linguistic Theory* **21**: 1–37.

Deane, Paul D. 1992. *Grammar in Mind and Brain*. Berlin/New York: Mouton de Gruyter.

Dekeyser, Xavier. 1990. Preposition stranding and relative complementizer deletion: Implicational tendencies in English and other Germanic languages. In: Sylvia Adamson, Vivian Low, Nigel Vincent and Susan Wright. *Papers from the 5th International Conference on English Historical Linguistics*. Amsterdam/ Philadelphia: Benjamins, 87–109.

Dekydtspotter, Laurent, Rex A. Sprouse and Bruce Anderson. 1998. Interlanguage A-bar dependencies: Binding construals, null prepositions and Universal Grammar. *Second Language Research* **14**: 341–58.

Denison, David. 1985. Why Old English had no prepositional passive. *English Studies* **3**: 189–204.

1993. *English Historical Syntax*. London/New York: Longman.

Department of Education and Science. 1989. *English for Ages 5 to 16*. ('The Cox Report'.) London: HMSO.

Diessel, Holger. 2006. Komplexe Konstruktionen im Erstspracherwerb. In: Fischer and Stefanowitsch, eds. 2006: 39–54.

Diessel, Holger and Michael Tomasello. 2000. The development of relative clauses in spontaneous child speech. *Cognitive Linguistics* **11**: 131–51.

Donohew, Grace. 1962. *A First Course in Luyia Grammar*. Nairobi: East African Literature Bureau.

Downing, Bruce T. 1978. Some universals of relative clause structure. In: Joseph H. Greenberg, ed. *Universals of Human Language*. Vol. IV: *Syntax*. Stanford, CA: Stanford University Press, 375–418.

Eckman, Fred. R. 2004. Universals, innateness and explanation in second language acquisition. *Studies in Language* **28**: 682–793.

Ellis, Nick C. 2003. Constructions, chunking, and connectionism: The emergence of second language structure. In: Catherine J. Doughty and Michael H. Long, eds. *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell.

Ellis, Rod. 1995. *Understanding Second Language Acquisition*. 10th edn. Oxford: Oxford University Press.

1998. *Second Language Acquisition*. 3rd edn. Oxford: Oxford University Press.

Ernst, Thomas. 2002. *The Syntax of Adjuncts*. Cambridge: Cambridge University Press.

Featherston, Sam. 2004. Bridge verbs and V2 verbs – the same thing in spades? *Zeitschrift für Sprachwissenschaft* **23**: 181–210.

2005. Magnitude estimation and what it can do for your syntax: Some wh-constraints in German. *Lingua* **115**: 1525–50.

Field, Andy. 2003. Repeated Measures ANOVA. www.statisticshell.com/repeated-measures.pdf.

2009. *Discovering Statistics Using SPSS*. London: Sage.

Fillmore, Charles J. 1977. The case for case reopened. In: Peter Cole and Jerrold M. Sadock, eds. *Syntax and Semantics*. Vol. VIII: *Grammatical Relations*. New York: Academic Press, 59–81.

1999. Inversion and constructional inheritance. In: Webelhuth, König and Kathol, eds. 1999: 113–28.

Fillmore, Charles J and Paul Kay. 1996. Construction Grammar. Unpublished Ms. University of California at Berkeley Department of Linguistics.

Filppula, Markku. 2004. Irish English: Morphology and syntax. In: Kortmann, Burridge, Mesthrie, Schneider and Upton, eds. 2004: 73–101.

Fischer, Kerstin and Anatol Stefanowitsch, eds. 2006. *Konstruktionsgrammatik: Von der Anwendung zur Theorie*. Tübingen: Stauffenberg.

Fischer, Olga. 1992. Syntax. In: Norman Blake, ed. 1992. *The Cambridge History of the English Language*. Vol. II: *1066–1476*. Cambridge: Cambridge University Press, 207–408.

Fischer, Olga, Ans van Kemenade, Willem Koopman and Wim van der Wurff. 2000. *The Syntax of Early English*. Cambridge: Cambridge University Press.

Fisher, Ann. 1750. *A New Grammar: Being the Most Easy Guide to Speaking and Writing the English Language Properly and Correctly*. Repr. in facs. by Robin C. Alston (1968): *English Linguistics* 130. Menston: The Scholar Press Limited.

Fleischer, Jürg. 2002. Preposition stranding in German dialects. In: Barbiers Sief, Leonie Cornips and Susanne van der Kleij, eds. *Syntactic Microvariation*. www.meertens.knaw.nl/books/synmic/pdf/fleischer.

Fowler, Henry W. 1926. *A Dictionary of Modern English Usage*. Oxford: Oxford University Press.

Freeman, Daniel H., Jr. 1987. *Applied Categorical Data Analysis*. New York: Marcel Dekker.

Gathumbi, Agnes, Eunice Kimaliro, Harriet Mugambi, Jane Bwonya, Julie Newman and Ngungui Kiio. 2003. *New Integrated English: Teacher's Guide 1*. Nairobi: The Jomo Kenyatta Foundation.

2004a. *New Integrated English: Teacher's Guide 2*. Nairobi: The Jomo Kenyatta Foundation.

2004b. *New Integrated English: Teacher's Guide 3*. Nairobi: The Jomo Kenyatta Foundation.

2005. *New Integrated English: Teacher's Guide 4*. Nairobi: The Jomo Kenyatta Foundation.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Gilquin, Gaëtanelle and Stefan Th. Gries. 2009. Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory* **5**: 1–26.

Ginzburg, Jonathan and Ivan A. Sag. 2000. *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives*. Stanford, CA: CSLI Publications.

Goldberg, Adele E. 2003. Constructions: A new theoretical approach to language. *TRENDS in Cognitive Sciences* **7**: 219–24.

2006. *Constructions at Work: The Nature of Generalisation in Language*. Oxford: Oxford University Press.

Görlach, Manfred. 1994. *Einführung ins Frühneuenglische.* 2nd edn. Heidelberg: Winter.

1999. *English in Nineteenth-Century England: An Introduction.* Cambridge: Cambridge University Press.

Granger, Sylviane. 1998. The computer learner corpus: A versatile new source of data for SLA research. In: Sylviane Granger, ed. *Learner English on Computer.* London: Longman, 3–18.

Gravetter, Frederick J. and Larry B. Wallnau. 1992. *Statistics for the Behavioral Sciences: A First Course for Students of Psychology and Education.* 3rd edn. New York et al: West Publishing Company.

Greenbaum, Sidney, ed. 1996. *Comparing English Worldwide: The International Corpus of English.* Oxford: Oxford University Press.

Grewendorf, Günther. 2002. *Minimalistische Syntax.* Tübingen/Basle: A. Franke Verlag.

Gries, Stefan Th. 2002. Preposition stranding in English: Predicting speakers' behaviour. In: Vida Samiian, ed. *Proceedings of the Western Conference on Linguistics.* Vol. XII. Fresno, CA: Department of Linguistics, California State University, 230–41.

2003. *Multifactorial Analysis in Corpus Linguistics.* New York/London: Continuum.

2004a. *Coll.analysis 3. A Program for R for Windows 2.x.* people.freenet.de/Stefan_Th_Gries.

2004b. *HCFA 3.2 – A Program for Hierarchical Configural Frequency Analysis for R for Windows.* people.freenet.de/Stefan_Th_Gries.

2008. *Statistik für Sprachwissenschaftler.* (Studienbuch zur Linguistik 13.) Göttingen: Vandenhoeck and Ruprecht.

Gries, Stefan Th., Beate Hampe and Doris Schönefeld. 2005. Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* **16**: 635–76.

Guy, Gregory R. and Robert Bayley. 1995. On the choice of relative pronouns in English. *American Speech* **70**: 148–62.

Haberzettl, Stefanie. 2006. Konstruktionen im Zweitspracherwerb. In: Fischer and Stefanowitsch, eds. 2006: 55–77.

Halliday, M. A. K. 1988. On the ineffability of grammatical categories. In: James Benson, Michael J. Cummings and William S. Greaves, eds. *Linguistics in a Systemic Perspective.* Amsterdam/Philadelphia: Benjamins, 27–51.

Hamilton, Robert. 1994. Sum: Null-prep and piping. Posting on the *Linguist List*, 28.06.1994. www.umich.edu/~archive/linguistics/linguist.list/volume.5/no.751–800.

Hancock, Ian F. and Rachel Angogo. 1982. English in East Africa. In: Richard W. Bailey and Manfred Görlach. *English as a World Language.* Ann Arbor: The University of Michigan Press.

Hauser, Marc D., Noam Chomsky and W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* **298**: 1569–79.

Hawkins, John A. 1994. *A Performance Theory of Order and Constituency.* Cambridge: Cambridge University Press.

1999. Processing complexity and filler-gap dependencies across grammars. *Language* **75**: 245–85.

2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hawkins, Roger. 2001. *Second Language Syntax: A Generative Introduction*. Oxford/ Malden, MA: Blackwell.

Hay, Jennifer B. and R. H. Baayen. 2005. Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences* **9**: 342–8.

Heine, Bernd. 1982. *Boni Dialects*. (Language and Dialect Atlas of Kenya 10.) Berlin: Dietrich Reimer.

Heine, Bernd and William J. G. Möhlig. 1980. *Geographical and Historical Introduction: Language and Society*. (Language and Dialect Atlas of Kenya 1.) Berlin: Dietrich Reimer.

Hoekstra, Jarich. 1995. Preposition stranding and resumptivity in West Germanic. In: Hubert Haider, Susan Olsen and Sten Vikner, eds. *Studies in Comparative Germanic Syntax*. Dordrecht: Kluwer, 95–118.

Hoffmann, Thomas. 2003. Preposition stranding vs pied-piping in British English relative clauses: A corpus-based study. Unpublished M.A. thesis, University of Regensburg.

2005. Variable vs categorical effects: Preposition pied piping and stranding in British English relative clauses. *Journal of English Linguistics* **33**: 257–97.

2006. Corpora and introspection as corroborating evidence: The case of preposition placement in English relative clauses. *Corpus Linguistics and Linguistic Theory* **2**: 165–95.

2007a. 'Good is good and bad is bad': But how do we know which one we had?. *Corpus Linguistics and Linguistic Theory* **3**: 87–98.

2007b. 'I need data which I can rely on': Corroborating empirical evidence on preposition placement in English relative clauses. In: Sam Featherston and Wolfgang Sternefeld, eds. *Roots: Linguistics in Search of its Evidential Base*. Berlin: Mouton de Gruyter, 161–83.

2008. English relative clauses and Construction Grammar: A topic which preposition placement can shed light on? In: Graeme Trousdale and Nicolas Gisborne, eds. *Constructional Explanations in English Grammar*. Berlin: Mouton de Gruyter, 77–112.

Hornstein, Norbert and Amy Weinberg. 1981. Case theory and preposition stranding. *Linguistic Inquiry* **12**: 55–91.

Hosmer, David W. and Stanley Lemeshow. 2000. *Applied Logistic Regression*. 2nd edn. New York: Wiley.

Huddleston, Rodney. 1984. *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.

2002a. The clause: Complements. In: Huddleston and Pullum, eds. 2002: 213–321.

2002b. Clause type and illocutionary force. In: Huddleston and Pullum, eds. 2002: 851–946.

2002c. Non-finite and verbless clauses. In: Huddleston and Pullum, eds. 2002: 1171–272.

Huddleston, Rodney and Geoffrey K. Pullum, eds. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Huddleston, Rodney, Geoffrey K. Pullum and Peter Peterson. 2002. Relative constructions and unbound dependencies. In: Huddleston and Pullum, eds. 2002: 1031–96.

Hudson, Richard A. 1996. *Sociolinguistics.* 2nd edn. Cambridge: Cambridge University Press.

Hudson-Ettle, Diana M. and Josef Schmied. 1999. Manual to accompany the East African Component of the International Corpus of English. Unpublished Ms. Chemnitz University of Technology.

Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford: Oxford University Press.

Johansson, Christine and Christer Geisler. 1998. Pied piping in spoken English. In: Antoinette Renouf, ed. *Explorations in Corpus Linguistics.* Amsterdam: Rodopi, 67–82.

Johnson, Daniel E. 2009a. Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass* **3**: 359–83.

2009b. RBRUL MANUAL. www.ling.upenn.edu/~johnson4/Rbrul_manual.html.

Johnson, Keith. 2008. *Quantitative Methods in Linguistics.* Oxford/Malden, MA: Blackwell.

Kanyoro, Musimbi R. A. 1991. The politics of the English language in Kenya and Tanzania. In: Cheshire, ed. 1991: 402–17.

Kao, Rong-Rong. 2001. Where have the prepositions gone? A study of English prepositional verbs and input enhancement in instructed SLA. *IRAL* **39**: 195–215.

Kay, Paul and Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalisations: The *What's X doing Y?* construction. *Language* **75**: 1–33.

Keller, Frank, Martin Corley, Steffan Corley, Lars Konienczny and Amalia Todirascu. 1998. WebExp: A Java toolbox for web-based psychological experiments. Technical Report HCRC/TR-99, Human Communication Research Centre, University of Edinburgh.

Keller, Frank and Theodora Alexopoulou. 2005. A crosslinguistic, experimental study of resumptive pronouns and *that*-trace effects. In: Bruno G. Bara, Lawrence Barsalou and Monica Bucciarelli, eds. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, 1120–5. www.cogsci.rpi.edu/CSJarchive/Proceedings/2005/docs/p1120.pdf.

Keller, Frank. 2000. Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality. Ph.D. thesis, University of Edinburgh.

Kempen, Gerard and Karin Harbusch. 2005. The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German clauses. In: Kepser and Reis, eds. 2005. 327–47.

Kepser, Stephan and Marga Reis, eds. 2005. *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives.* Berlin/New York: Mouton de Gruyter.

Kioko, Angelina Nduku and Margaret Jepkirui Muthwii. 2004. English variety for the public domain in Kenya: Speaker's attitudes and views. In: Margaret Jepkirui Muthwii and Angelina Nduku Kioko, eds. *New Language Bearings in Africa: A Fresh Quest.* Clevedon: Multilingual Matters Ltd., 34–58.

Kjellmer, Göran. 1988. 'What a night on which to die': On symmetry in English relative clauses. *English Studies* **6**: 559–68.

Klein, Elaine C. 2001. (Mis)constructing null prepositions in L2 intergrammars: A commentary and a proposal. *Second Language Research* **17**: 37–70.

Kline, Paul. 1993. *The Handbook of Psychological Testing.* London/New York: Routledge.

Kluender, Robert. 1990. A neurophysiological investigation of Wh-Islands. In: Kira Hall *et al.*, eds. *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society.* Berkeley: Berkeley Linguistics Society, 187–204.

König, Ekkehard and Bernd Kortmann. 1991. On the reanalysis of verbs as prepositions. In: Gisa Rauh, ed. *Approaches to Prepositions.* (Tübinger Beiträge zur Linguistik 358.) Tübingen: Gunter Narr, 109–25.

Kortmann, Bernd, Kate Burridge, Rajend Mesthrie, Edgar W. Schneider and Clive Upton, eds. 2004. *A Handbook of Varieties of English.* Vol. II: *Morphology and Syntax.* Berlin and New York: Mouton de Gruyter.

Krifka, Manfred. 1995. Swahili. In: Joachim Jacobs, Arnim von Stechow, Wolfgang Strenefeld and Theo Vennemann, eds. *Syntax: An International Handbook of Contemporary Research.* Vol. II. Berlin/New York: Walter de Gruyter, 1397–418.

Kubota, Yusuke. 2003. Yet another HPSG-analysis for free relative clauses in German. In: Jong-Bok Kim and Stephan Wechsler, eds. *Proceedings of the 9th International Conference on HPSG.* Stanford, CA: Stanford University, 147–67.

Labov, William. 1966. *The Social Stratification of English in New York City.* Washington, DC: Centre for Applied Linguistics.

1972. *Sociolinguistic Patterns.* Philadelphia: University of Philadelphia Press.

1973. The boundaries of words and their meaning. In: Charles-James Bailey and Roger W. Shuy, eds. *New Ways of Analyzing Variation in English.* Washington, DC: Georgetown University Press, 340–73.

1975. Empirical foundations of linguistic theory. In: Robert P. 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.* Lisse: Peter de Ridder, 77–133.

Lakoff, George. 1987. *Women, Fire and Dangerous Things: What Categories Reveal About the Mind.* Chicago/London: The University of Chicago Press.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar.* Vol. I: *Theoretical Prerequisites.* Stanford, CA: Stanford University Press.

2005. Construction Grammars: Cognitive, radical, and less so. In: Francisco J. Ruiz de Mendoza Ibánez and M. Sandra Pena Cervel, eds. *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction.* Berlin/New York: Mouton de Gruyter, 101–59.

Larson, Richard K. 1988. Implicit arguments in situation semantics. *Linguistics and Philosophy* **11**: 169–201.

Leaky, Louis S. B. 1959. *First Lessons in Kikuyu.* Dar es Salaam: Eagle Press.

Leech, Geoffrey. 1992. Corpus and theories of linguistic performance. In: Jan Svartvik, ed. *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991.* Berlin and New York: Mouton de Gruyter, 105–22.

1996. *An A–Z of English Grammar and Usage.* Harlow, Essex: Longman.

1998. Preface. In: Sylviane Granger, ed. *Learner English on Computer*. London: Longman, xiv–xx.
2007. New resources, or just better old ones? The Holy Grail of representativeness. In: Marianne Hundt, Nadja Nesselhauf and Carolin Biewer, eds. *Corpus Linguistics and the Web*. Amsterdam/New York: Rodopi, 133–49.
Leisi, Ernst. 1985. *Das heutige Englisch: Wesenszüge und Probleme*. 7th edn. Heidelberg: Carl Winter Universitätsverlag.
Lieven, Elena, Heike Behrens, Jennifer Speares and Michael Tomasello. 2003. Early syntactic creativity: A usage based approach. *Journal of Child Language* **30**: 333–70.
Lemnitzer, Lothar and Heike Zinsmeister. 2006. *Korpuslinguistik: Eine Einführung*. Tübingen: Gunter Narr.
Lowth, Robert. 1762. *A Short Introduction to English Grammar*. 2nd edn. Repr. in facs. by Robin C. Alston (1967): *English Linguistics* 18. Menston: The Scholar Press Limited.
Lu, Bingfu. 2002. How does language encode performance limitation into its structure. www.people.fas.harvard.edu/~whu/China/chunk.doc.
Mackey, Alison. 2006. Second language acquisition. In: Ralph W. Fasold and Jeff Connor-Linton, eds. *An Introduction to Language and Linguistics*. Cambridge: Cambridge University Press, 433–63.
Maindonald, John and John Braun. 2003. *Data Analysis and Graphics Using R: An Example-based Approach*. Cambridge: Cambridge University Press.
Mazurkewich, Irene. 1985. Syntactic Markedness and language acquisition. *Studies in Second Language Acquisition* **7**: 15–36.
Mbaabu, Ireri. 1996. *Language Policy in East Africa: A Dependency Theory Perspective*. Nairobi: Educational Research Publications.
McDaniel, Dana and Wayne Cowart. 1999. Experimental evidence for a Minimalist account of English resumptive pronouns. *Cognition* **70**: 15–24.
McDaniel, Dana, Cecile McKee and Judy B. Bernstein. 1998. How children's relatives solve a problem for Minimalism. *Language* **74**: 308–34.
McEnery, Tony and Andrew Wilson. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
McKercher, David A. 1996. On the syntax and semantics of English adverbials. M.A. thesis, University of Victoria. individual.utoronto.ca/McKercher-MA-thesis.pdf.
Mesthrie, Rajend. 2004. Black South African English: Morphology and syntax. In: Kortmann, Burridge, Mesthrie, Schneider and Upton, eds. 2004: 962–73.
Mesthrie, Rajend and Rakesh M. Bhatt. 2008. *World Englishes: The Study of New Linguistic Varieties*. Cambridge: Cambridge University Press.
Miller, Jim. 2004. Scottish English: Morphology and syntax. In: Kortmann, Burridge, Mesthrie, Schneider and Upton, eds. 2004: 47–72.
Milroy, Lesley and Matthew Gordon. 2003. *Sociolinguistics: Method and Interpretation*. Malden, MA: Blackwell.
Montgomery, Michael. 1971. *Effective English: Teacher's Book 2*. London: Evans Brothers.
Mufwene, Salikoko S. 2004. Gullah: Morphology and syntax. In: Kortmann, Burridge, Mesthrie, Schneider and Upton, eds. 2004: 356–73.

Müller, Stefan. 1999. An HPSG-Analysis for free relative clauses in German. *Grammars* **2**: 53–105.

2006. Resultativkonstruktionen, Partikelverben und syntaktische vs lexikonbasierte Konstruktionen. In: Fischer and Stefanowitsch, eds. 2006: 176–202.

Murray, Lindley. 1795. *English Grammar: Adapted to the Different Classes of Learners*. Repr. in facs. by Robin C. Alston (1968): *English Linguistics* 106. Menston. The Scholar Press Limited.

Musau, Paul M. 2004. Linguistic human rights in Africa: Challenges and prospects for indigenous languages in Kenya. In: Margaret Jepkirui Muthwii and Angelina Nduku Kioko, eds. *New Language Bearings in Africa: A Fresh Quest*. Clevedon: Multilingual Matters Ltd., 59–68.

Mwangi, Serah. 2004. Preposition vanishing in Kenya: A case of syntactic simplification in a variety of East African English. *English Today* **77**: 27–32.

Nelson, Gerald, Sean Wallis and Bas Aarts. 2002. *Exploring Natural Language: Working with the British Component of the International Corpus of English*. Amsterdam/Philadelphia: Benjamins.

Olofsson, Arne. 1981. *Relative Junctions in Written American English*. Göteborg: ACTA Universitatis Gothoburgensis.

Omondi, Lucia Ndong'a. 1982. *The Major Syntactic Structures of Dholuo*. (Language and Dialect Atlas of Kenya Supplement 1.) Berlin: Dietrich Riemer Verlag.

Palmer, Harold E. 1974. *The Teaching of Oral English*. 2nd edn. London: Longman.

Paolillo, John C. 2002. *Analyzing Linguistic Variation: Statistical Models and Methods*. Stanford, CA: CSLI Publications.

Parkinson, Tom, Matt Philips and Will Gourlay. 2006. *Lonely Planet: Kenya*. 6th edn. London: Lonely Planet Publication Pty Ltd.

Patrick, Peter L. 2004. Jamaican Creole: Morphology and syntax. In: Kortmann, Burridge, Mesthrie, Schneider and Upton, eds. 2004: 407–38.

Payne, John and Rodney Huddleston. 2002. Nouns and noun phrases. In: Huddleston and Pullum, eds. 2002: 323–524.

Penke, Martina and Anette Rosenbach. 2004. What counts as evidence in linguistics? An introduction. *Studies in Language* **28**: 480–526.

Pesetsky, David. 1998. Some Optimality principles of sentence production. In: Pilar Barbosa *et al.*, eds. *Is the best Good Enough? Optimality and Competition in Syntax*. Cambridge, MA: The MIT Press, 337–83.

Philp, A. M. 2001. English grammar in British schools. In: Rajend Mesthrie, ed. *Concise Encyclopedia of Sociolinguistics*. Amsterdam: Elsevier, 723–35.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In: Bybee and Hopper, eds. 2001: 137–57.

Pollard, Carl and Ivan A. Sag. 1987. *Information-based Syntax and Semantics*. Vol. I: *Fundamentals*. Stanford, CA: CSLI Publications.

1994. *Head-driven Phrase Structure Grammar*. Chicago/London: The University of Chicago Press.

Preston, Dennis R. 1996. Variationist perspectives on second language acquisition. In: Bayley and Preston, eds. 1996: 1–45.

Pullum, Geoffrey K. and Barbara C. Scholz. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* **19**: 9–50.

Pullum, Geoffrey K. and Rodney Huddleston. 2002. Prepositions and prepositional phrases. In: Huddleston and Pullum, eds. 2002: 597–661.

Quirk, Randolph. 1957. Relative clauses in educated spoken English. *English Studies* **38**: 97–109.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. www.R-project.org.

Radford, Andrew. 1981. *Transformational Syntax: A Student's Guide to Chomsky's Extended Standard Theory*. Cambridge: Cambridge University Press.

1988. *Transformational Grammar: A First Course*. Cambridge: Cambridge University Press.

1997. *Syntax: A Minimalist Introduction*. Cambridge: Cambridge University Press.

2004. *Minimalist Syntax: Exploring the Structure of English*. Cambridge: Cambridge University Press.

Rand, Eric. 1969. *Constructing Sentences*. New York: Holt, Rinehart and Winston Inc.

Rezai, Mohammad Javad. 2006. Preposition stranding and pied piping in second language acquisition. *Essex Graduate Student Papers in Language and Linguistics* 8. www.sx.ac.uk/linguistics/pgr/egspll/volume8/PDFs/M.%20Javad%20Rezai.pdf.

Robinson, John S., Helen. R. Lawrence and Sali A. Tagliamonte. 2001. GOLDVARB 2001: A multivariate analysis application for Windows. www.york.ac.uk/depts/lang/webstuff/goldvarb.

Rosch, Eleanor. 1978. Principles of categorization. In: Eleanor Rosch and Barbara B. Loyd, eds. *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, 27–48.

Ross, John R. 1986. *Infinite Syntax!* Norwood, NJ: Ablex Publishing Corporation.

Rumelhart, David E., James L. McClelland and the PDP research group. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. 2 vols. Cambridge, MA: MIT Press.

Sadighi, F., M. R. Parhizgar and M. Saadat. 2004. Preposition pied-piping and preposition stranding constructions in the interlanguage of Iranian EFL learners. *Asian EFL Online Journal* **6**: Article 3. www.asian-efl-journal.com.

Saffran, Jenny R. 2001. The use of predictive dependencies in language learning. *Journal of Memory and Language* **44**: 493–515.

2002. Constraints on statistical language learning. *Journal of Memory and Language* **47**: 172–96.

Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* **33**: 431–84.

2007. Sign-Based Construction Grammar. An informal synopsis. Unpublished Ms. lingo.stanford.edu/sag/papers/theo-syno.pdf.

Sampson, Geoffrey. 2001. *Empirical Linguistics*. London/New York: Continuum.

Schmied, Josef. 1991a. Relative constructions in British and Indian English. Postdoctoral diss., University of Bayreuth.

1991b. *English in Africa: An Introduction*. London: Longman.

1996. Second-language corpora. In Greenbaum, ed. 1996: 182–96.

2004a. East African English phonology. In: Bernd Kortmann, Kate Burridge, Rajend Mesthrie, Edgar W. Schneider and Clive Upton, eds. *A Handbook of Varieties of English*. Vol. I: *Phonology*. Berlin/New York: Mouton de Gruyter, 918–30.

2004b. East African English (Kenya, Uganda, Tanzania): Morphology and Syntax. In: Kortmann, Burridge, Mesthrie, Schneider and Upton, eds. 2004: 929–47.

Schneider, Edgar W. 1992. *Who*(*m*)? Constraints on the loss of case marking of wh-pronouns in the English of Shakespeare and other poets of the Early Modern English period. In: Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen and Irma Taavitsainen, eds. *History of Englishes: New Methods and Interpretations in Historical Linguistics*. Berlin/New York: Peter Lang, 437–52.

2007. *Postcolonical English: Varieties around the World*. Cambridge: Cambridge University Press.

Schütze, Carson T. 1996. *The Empirical Base of Linguistics: Grammaticality Judgements and Linguistic Methodology*. Chicago: Chicago University Press.

Sigley, Robert J. 1997. Choosing your relatives: Relative clauses in New Zealand English. Ph.D. thesis, Victoria University of Wellington.

2003. The importance of interaction effects. *Language Variation and Change* **15**: 227–53.

Skandera, Paul. 2003. *Drawing a Map of Africa: Idiom in Kenyan English*. Tübingen: Gunter Narr.

Sorace, Antonella and Frank Keller. 2005. Gradience in linguistic data. *Lingua* **115**: 1497–525.

Stefanowitsch, Anatol and Stefan Th. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* **1**: 1–43.

Steinberg, Danny D. and Natalia V. Sciarini. 2006. *An Introduction to Psycholinguistics*. 2nd edn. London: Longman.

Sundby, Bertil, Anne Kari Bjørge and Kari E. Haugland. 1991. *A Dictionary of English Normative Grammar 1700–1800*. Amsterdam/Philadelphia: Benjamins.

Szmrecsanyi, Benedikt. 2006. Review of Rens Bod, Jennifer Hay, and Stefanie Jannedy eds. 2003. *Probabilistic Linguistics*. Cambridge, MA: MIT Press. *Corpus Linguistics and Linguistic Theory* 2: 79–94.

Taylor, John. 1995. *Linguistic Categorization: Prototypes in Linguistic Theory*. 2nd edn. Oxford: Oxford University Press.

Tieken-Boon van Ostade, Ingrid. 2006. Eighteenth-century Prescriptivism and the Norms of Correctness. In: Ans van Kemenade and Bettelou Los, eds. *The Handbook of the History of English*. Malden, MA: Blackwell, 539–57.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

2006. Konstruktionsgrammatik und Erstspracherwerb. In: Fischer and Stefanowitsch, eds. 2006: 19–37.

Traugott, Elizabeth C. 1992. Syntax. In: Richard M. Hogg, ed. *The Cambridge History of the English Language*. Vol. I: *The Beginnings to 1066*. Cambridge: Cambridge University Press, 168–289.

Trotta, Joe. 2000. *Wh-clauses in English: Aspects of Theory and Description*. Amsterdam/Philadelphia, GA: Rodopi.

Trudgill, Peter. 1995. *Sociolinguistics: An Introduction to Language and Society*. London: Penguin.

Tseng, Jesse. 2006. English prepositional passives in HPSG. In: Shuly Winter, ed. *Proceedings of the FG-2006 Conference*. Stanford, CA: CSLI Publications, 171–183. csli-publications.stanford.edu.

Ungerer, Friedrich, Gerhard E. H. Meier, Klaus Schäfer and Shirley B. Lechler. 1996. *A Grammar of Present-Day English*. Stuttgart: Klett.

Van der Auwera, Johan. 1985. Relative *that* – a centennial dispute. *Journal of Linguistics* **21**: 149–79.

Van den Eynden, Nadine. 1996. Aspects of prepositional placement in English. In: Juhani Klemola, Merja Kytö and Matti Rissanen, eds. *Speech Past and Present: Studies in English Dialectology in Memory of Ossi Ihalainen*. Frankfurt am Main: Peter Lang, 426–46.

Van Eynde, F. 2004. Pied Piping is a local dependency. In: Stefan Müller, ed. *Proceedings of the HPSG04 Conference*. Stanford, CA: CSLI Publications. csli-publications.stanford.edu.

Van Riemsdijk, Henk. 2000. Free relatives. *Syncom: Case 44*. faculty.washington.edu/bcitko/LING580/van Riemsdijk_syncom.pdf.

Visser, F. Th. 1963. *An Historical Syntax of the English Language*. Vol I: *Syntactical Units with one Verb*. Leiden: Brill.

Vitale, Anthony J. 1981. *Swahili Syntax*. Dordrecht/NL, Cinnaminson/USA: Foris Publications.

Ward, Gregory, Betty Birner and Rodney Huddleston. 2002. Information packaging. In: Huddleston and Pullum, eds. 2002: 1363–448.

Webelhuth, Gert, Jean-Pierre König and Andreas Kathol, eds. 1999. *Lexical and Constructional Aspects of Linguistic Explanation*. Stanford, CA: Berkeley Linguistic Society.

Weiner, Edmund S. C. 1983. *The Oxford Guide to English Usage*. Oxford: Clarendon Press.

Weiß, Helmut. 2001. On two types of natural languages: Some consequences for linguistics. *Theoretical Linguistics* **27**: 87–103.

White, Lydia. 2003. *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Whitely, Wilfred H. and M. G. Muli. 1962. *Practical Introduction to Kamba*. London: Oxford University Press.

Wilcock, Graham. 1999. Lexicalization of context. In: Webelhuth, König and Kathol, eds. 1999: 373–87.

Wittengenstein, Ludwig. 1953. *Philosophical Investigations*. Translated by Gertrude E. M. Anscombe. Oxford: Blackwell.

Woods, Anthony, Paul Fletcher and Arthur Hughes. 1986. *Statistics in Language Studies*. Cambridge: Cambridge University Press.

Yáñez-Bouza, Nuria. 2004. Preposition stranding in Early Modern English: A corpus-based approach. Paper presented at the *13th International Conference of English Historical Linguistics* (*ICEHL*), University of Vienna, Austria. 23–27 August.

2006. Prescriptivism and preposition stranding in eighteenth-century prose. *Historical-Socio-Linguistics/Socio-Historical-Linguistics* 6. www.let.leidenuni.nl/hsl_shl/preposition%20stranding.htm.

Young, Richard and Robert Bayley. 1996. VABRUL analysis for second language acquisition research. In: Bayley and Preston, eds. 1996: 253–306.

Index