

Cambridge University Press

978-0-521-76047-8 - Preposition Placement in English: A Usage-based Approach

Thomas Hoffmann

Excerpt

[More information](#)

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. Hoffmann 2006; 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 Hamelin who followed the pied piper in the well-known fairy tale (Ross 1986: 126, n. 23).

Cambridge University Press

978-0-521-76047-8 - Preposition Placement in English: A Usage-based Approach

Thomas Hoffmann

Excerpt

[More information](#)

2 Introduction

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.

Cambridge University Press

978-0-521-76047-8 - Preposition Placement in English: A Usage-based Approach

Thomas Hoffmann

Excerpt

[More information](#)**1.2 World Englishes and usage-based linguistics 3**

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

4 Introduction

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

Cambridge University Press

978-0-521-76047-8 - Preposition Placement in English: A Usage-based Approach

Thomas Hoffmann

Excerpt

[More information](#)

1.2 World Englishes and usage-based linguistics 5

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*-TENSE² *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.

Cambridge University Press

978-0-521-76047-8 - Preposition Placement in English: A Usage-based Approach

Thomas Hoffmann

Excerpt

[More information](#)

6 Introduction

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).

Cambridge University Press

978-0-521-76047-8 - Preposition Placement in English: A Usage-based Approach

Thomas Hoffmann

Excerpt

[More information](#)**1.3 Outline 7**

(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

Cambridge University Press

978-0-521-76047-8 - Preposition Placement in English: A Usage-based Approach

Thomas Hoffmann

Excerpt

[More information](#)

8 Introduction

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’

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).

10 Corroborating evidence: Data and methodology

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