The co-evolution of syntactic and pragmatic complexity: diachronic and cross-linguistic aspects of pseudoclefts

Christian Koops
Rice University
ckoops@rice.edu

Martin Hilpert
Freiburg Institute for Advanced Studies
martin.hilpert@frias.uni-freiburg.de

This chapter examines the diachronic rise of a syntactically and pragmatically complex construction type: pseudoclefts. Given that cleft constructions combine available components of grammar – relative clauses and copular clauses – do they arise in full-fledged form? If they emerge gradually, what constrains their development? We first present a corpus-based analysis of the history of English pseudoclefts and develop qualitative and quantitative measures to identify properties of pseudoclefts at different developmental stages. We then apply the same measures of grammaticalization in a synchronic comparison of pseudoclefts in contemporary spoken and written German, Swedish, and English in order to test their cross-linguistic validity. We find that pseudoclefts develop gradually in a process driven by the pragmatic exploitation of their presuppositional structure (Lambrecht 1994).[1]

1. Introduction: Cleft constructions and grammaticalization

The diachrony of information structure constructions has been among the foundational issues of functional typological linguistics as well as research on grammaticalization. The synchronic finding that morphological focus markers often resemble copulas and that the non-focused (presupposed) part of focus constructions often exhibits properties of relative clauses allowed the reconstruction of a diachronic process in which morphological focus marking systems develop from syntactic (especially cleft) constructions (Givón 1979, Heine & Reh 1984). This cleft-to-focus marker pathway, as an example of the simplification of bi-clausal to monoclausal syntax, has been the phenomenon of primary interest thus far. Here we take a different approach to the topic. Rather than viewing focus constructions from the perspective of the reduction of syntax to morphology, we examine the emergence of cleft constructions themselves.

The reason for the small amount of attention paid to the diachrony of clefts may lie in a view of cleft constructions as, in the words of Harris and Campbell (1995: 54), "universally available syntactic constructions."

> While not every language has clefts, it is likely that such constructions are easily added to grammars. The focus cleft may be so widely available because it is structurally equivalent to a copular clause with a relative clause modifying one of its constituents (56).

If clefts are structurally equivalent to their diachronic sources – a copular clause and a relative clause – it would seem that they can reveal little, if anything, about the rise of grammatical complexity, as there is no need for a development to take place. In fact, they might be expected to arise more or less instantaneously and in full-fledged form as soon as speakers decide to re-arrange existing grammatical resources for a new purpose.

Harris and Campbell suggest that even universally available constructions like clefts undergo some development, however. They start out as "exploratory" expressions that may "catch on." Crucially, "[o]nly when the expression is used *in additional contexts* and is *generalized* ... may we speak of a grammatical change having taken place." (54, emphasis added) It remains unclear what characterizes the exploratory phase in the case of cleft constructions. Also, what is the original context of use and what are its extensions? And what motivates the process of generalization that results in a completely productive grammatical construction?

In this paper, we examine in detail one type of cleft construction, pseudoclefts, in order to address these questions. Our aim is to develop a grammaticalization scenario that allows less fully developed pseudoclefts to be distinguished from more grammaticalized ones. We first present a quantitative, text-based analysis of the 300-year history of English *what*-clefts in order to identify relevant properties of diachronically earlier and later instances of this construction (Section 2). Next, we apply the findings of the historical analysis to synchronic, cross-linguistic data in another series of corpus analyses. We compare the present-day English pseudoclefts with those of German and Swedish, and show that the synchronic properties of German and Swedish pseudoclefts coincide well with those seen at earlier stages of English (Section 3). Finally, we interpret our findings in terms of Lambrecht's (1994) notions of *presuppositional structure* and its *pragmatic accommodation* (Sections 4 and 5).

1.1   *Form and function of pseudoclefts*

Clefts are traditionally understood as information structure constructions used to pragmatically structure a proposition into two parts: a presupposition and a focus (Prince 1978, Lambrecht 2001 *inter alia*). Our working definition of a cleft construction, following Lambrecht, is

> a complex sentence structure consisting of a matrix clause headed by a copula and a relative or relative-like clause whose relativized argument is coindexed with the predicative argument of the copula. Taken together, the matrix and the relative express a logically simple proposition, which can also be expressed in the form of a single clause without a change in truth conditions (2001: 467).

Besides the syntactically unmarked sentence structure in (1), speakers of present-day English have two focus-initial clefts and one focus-final one at their disposal, as illustrated by Lambrecht's examples in (2).

(1)   I like CHAMPAGNE.
(2) a. It is CHAMPAGNE (that) I like.   (*it*-cleft)
     b. What I like is CHAMPAGNE.   (*wh*-cleft)
     c. CHAMPAGNE is what I like.   (reverse *wh*-cleft)

Our concern in this chapter is the *wh*-cleft construction in (2b). We follow common usage in referring to it as *pseudocleft* and to the initial constituent as *wh-clause* even though, strictly speaking, we only deal with the most common type, *wh*-clefts built on *what*. The *wh*-clause expresses a presupposed open proposition, i.e. a proposition with a missing argument (e.g., 'I like x'), and the focus phrase provides a value for the variable in the presupposed open proposition (e.g., 'champagne'). The function of the construction is to specify the content of the focus phrase as the value for the variable contained in the *wh*-clause (e.g., x = 'champagne')

Their specificational function, i.e. the fact that they specify the value of a variable, distinguishes pseudoclefts like (3) from superficially similar predicational copular constructions like (4).

(3)   What he bought was champagne.
(4)   What he bought was expensive.   (Lambrecht 2001: 494)

The referentiality of the *wh*-clause distinguishes (4) from (3). As Lambrecht points out (2001: 494), whereas *what he bought* in (4) refers to a specific, previously known object or set of objects, the *wh*-clause in (3) is a (non-referential) propositional function. The point of uttering (4) is not to characterize a referent but to inform the hearer of what was bought.

While the English *it*-cleft construction has received some attention from a historical perspective (e.g., Ball 1991, 1994), the only previous study of the diachrony of pseudoclefts in English that we are aware of is Traugott's (2008) recent exploratory work on three pseudocleft types: *all-clefts*, *what*-clefts, and reverse *what*-clefts. Using data from a drama corpus, Traugott tests the specific hypothesis, derived from the work of Kim (1995) and Hopper (2001), that these constructions first emerged in particular interactional contexts. One of her findings is that while such a development may have occurred in the history of *all*-clefts, the same case cannot be made for *what*-clefts.

2.   The diachronic development of English pseudoclefts

We begin by tracing the development of pseudoclefts in English from their first appearance in texts at the end of the 17th century to their status in late 20th-century written and spoken discourse. Our data come from five corpora. The first two are the *Penn-Helsinki Parsed Corpus of Early Modern English* (PPCEME, Kroch et al. 2004) and the *Corpus of Late Modern English Texts* (CLMET, De Smet 2005). These are each composed of three subcorpora containing texts from consecutive 70-year periods

of Early and Late Modern English, respectively. The second and third corpora represent mid and late 20th-century written British English: the *Lancaster-Oslo/Bergen Corpus* and the *Freiburg-LOB corpus* (LOB and FLOB, Hofland et al. 1999). They contain matched quantities of text samples from identical discourse genres, published in 1961 and 1991, respectively. Finally, we present data from the *Santa Barbara Corpus of Spoken American English* (SBCSAE, Du Bois et al. 2000-2005), which consists of naturally occurring conversations recorded primarily in the 1980s.

To ensure that the pseudocleft tokens in each corpus were exhaustively identified, we performed a complete extraction of all instances of the word *what* (or *whate*, in the earliest periods), and then manually inspected the (thousands of) hits. This methodology not only ensured that no pseudoclefts were overlooked, but also made it possible to retrieve and quantify relevant constructional variants and other related constructions, as discussed in the following sections.

2.1   Discourse frequency

Figure 1 shows the basic text frequency of pseudoclefts across seven of the historical subperiods spanned by our first four corpora. With the exception of the unexpected spike in the 18th century, Figure 1 shows a gradual, sustained frequency increase, which continues through the 20th century. Its trajectory suggests that the underlying development is still ongoing, i.e. that the use of the construction continues to expand.

Figure 1. Basic text frequency (number of occurrences per 1 million words)

A basic frequency count as shown in Figure 1 may of course include unwanted genre effects. The rate of occurrence of pseudoclefts in a particular corpus may, for example, be greater simply because the contained texts are of a genre which for one reason or another favors its use. The risk of genre effects is particularly high in the earlier periods, which are represented by only a few dozen texts. We therefore performed a second, relative frequency analysis designed to neutralize the variable of genre. In this analysis, the construction's rate of occurrence is not related to the total number of words in the corpus (as in Figure 1) but to the frequency of a related and

(presumably) stable construction. We were able to take advantage of the fact that pseudoclefts, like the early example in (5), were predated by predicational copular constructions with referential *wh*-clause subjects, as in (6) (cf. (3) and (4), see also Traugott 2008).

(5) But what I wonder at is this: I find I did not start at his Proposal, as when it came from one whom I contemn'd. (specificational, i.e. pseudocleft)

(6) But what they do is so much above my understanding, I can't pretend to give an account of it. (predicational)

(PPCEME 3)

To the extent that both constructions are subject to the same or similar genre effects, such effects should thereby be neutralized. The result of the relative frequency analysis are presented in Figure 2. It shows that at the time pseudoclefts are first attested they were used four times less frequently than predicational copular constructions with *wh*-clause subjects. By the 20th century their ratio is reversed. Again, the data suggest a continuous frequency increase, except that unlike in Figure 1 no clear difference between the mid and late 20th century is discernable.

Figure 2. Relative frequency of pseudoclefts and predicational copular constructions with *wh*-clause subjects
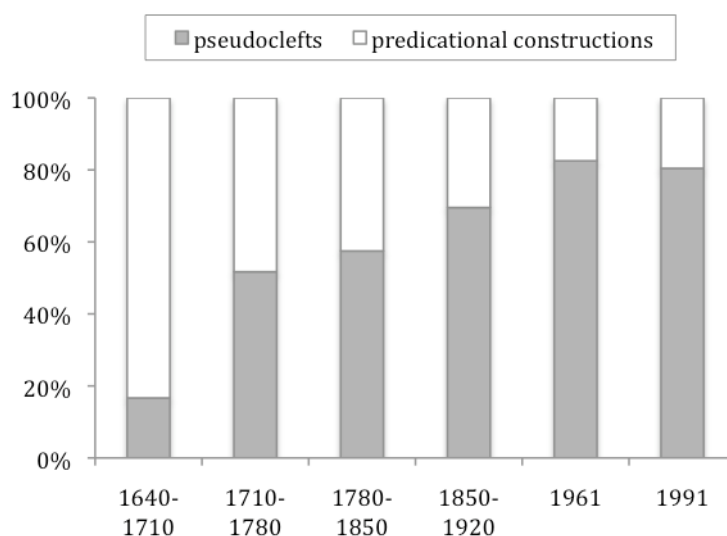
In summary, two frequency measures point to the same general picture: a sustained growth, suggesting that the development of English pseudoclefts has been progressing steadily over the past 300 years. Their rise in frequency suggests a continued expansion to new contexts of use.

2.2 Early pseudocleft variants

Before proceeding, we need to clarify the status of a non-canonical type of pseudocleft that we included in our count. From the time of their first appearance,

pseudoclefts like example (5) above were used alongside a kind of "copula-less" variant.

> (7)   And what rendered it the more dreary when we passed, there was a thick fog that hindered us from seeing above twenty yards from the carriage.
>
> (8)   In truth, I can find no excuse for you, and, what is more, I am certain you can find none for yourself.
>
> <div align="right">(CLMET 1)</div>

Examples (7) and (8) show preposed *what*-relatives whose relativized argument is co-referential with the following clause. Such cases differ from canonical pseudoclefts in their lack of syntactic integration, i.e. the absense of a copula and the fact that the clausal antecedent does not have the form of a syntactic argument. The only link between it and the *wh*-clause is their co-reference relation. Strictly speaking, then, these constructions fall outside of our initial definition of clefts.

Our decision to analyze such cases as pseudoclefts is based on their functional near-equivalence to canonical pseudoclefts. Crucially, they share the specificational function. It should also be pointed out that non-integrated pseudoclefts, including cases of copula "omission," are well-attested in spoken English (Weinert and Miller 1996, Hopper 2001, Koops and Ross-Hagebaum 2008). An example is given in (9).

> (9)   What they did, they took the stubs and they cleaned them up.
>
> <div align="right">(SBCSAE)</div>

The variant without a copula is even more common in spoken German (Günthner 2007, see also below). German also has both variants, and as in English the canonical variant is preferred in writing.

"Copula-less" pseudoclefts like (7) and (8) were most frequent in the 18th century, when they made up almost half of all attested tokens. Since then, they have gradually disappeared from texts. The two 20th-century corpora include only the formulaic *what is more* and *what is worse*. Their brief popularity and subsequent decline accounts for the frequency bump in the 1710-1780 period in Figure 1.

These facts suggest that as the canonical, "copula-full" pseudocleft gained in currency, cases like (7) came to be perceived as non-standard by comparison. Except for highly conventionalized phrases like *what is more* they were henceforth avoided in writing. Perhaps the lack of an overt syntactic link between the *wh*-clause and the focused clause made them appear too colloquial. In this respect, their status could be compared to that of left-dislocation, which is also largely banned from written genres despite its frequent use in spoken discourse.

2.3   *Modification of the wh-clause*

Free-standing *what*-relatives like those in (7) and (8) almost invariably expressed some form of comparison. Typically, the second of two conjuncts was characterized as having more or equally as much of some property as the first one. The same general phenomenon can be seen in the *wh*-clauses of pseudoclefts overall. They often contained an adverb of comparison.

(10)   But what <u>most</u> alarmed him was a hint that it was in her (Miss
        Matthews's) power to make Amelia as miserable as herself.

<div align="right">(CLMET 1)</div>

Another common form of modification were adverbials expressing an exceptional degree of some property.

(11)   But what <u>particularly</u> engaged our attention was a sealed note,
        superscribed, 'The copy of a letter to be sent to the two ladies at
        Thornhill-castle.'

<div align="right">(CLMET 1)</div>

In fact, early pseudoclefts only rarely occurred without some additional modification of the *wh*-clause. The attested types of modifiers are summarized in Table 1.

<div align="center">Table 1. Types of modification of the wh-clause</div>

Comparison	*What concerned her <u>even more</u> / <u>equally</u> ...*
Exceptional degree	*What is <u>especially</u> / <u>particularly</u> / <u>very</u> clear ...*
Deixis	*What is crucial <u>here</u> ... what he needed <u>now</u> ...*
Temporal adverbs	*What <u>finally</u> convinced him ... what <u>first</u> struck me ...*
Anaphoric *so*	*What made it <u>so</u> difficult ...*
Epistemic modality	*What I <u>actually</u> wanted ... what <u>may</u> appear odd ...*
Addition	*What she could <u>also</u> do ... what is notable <u>too</u> ...*

The uniting feature of the various expressions in Table 1 is that they implicitly or explicitly link the *wh*-clause proposition, and by extension the focused constituent, to the preceding discourse. For example, the *wh*-clause *what most alarmed him* in (10) implicitly makes reference to other causes for alarm. Similarly, the choice of the adverb *particularly* in (11) indexes other noticeable events previously mentioned in the narrative. Speaking more generally, these modifiers show that a prominent function of the early pseudoclefts was that of establishing topical coherence.

In this respect, their use did not differ very much from the way pseudoclefts are used today. In texts, pseudoclefts often occur at transition points as writers turn from one idea to the next (Jones and Jones 1985). In conversation, they are often used to re-orient the course of the ongoing talk, for example when resuming a temporarily suspended topic after a digression (Kim 1995). One major point of difference between pseudoclefts then and now, however, is the degree to which the link between the prior discourse and the focused constituent was routinely made explicit, as seen by the near-obligatoriness of some additional modification used to this effect.

The rate of occurrence of the modifiers in Table 1 provides a measure of the diachronic generalization of the construction. Figure 3 shows how the proportion of modified and unmodified *wh*-clauses in pseudoclefts shifted. (In order to base the measure on large enough numbers of examples, only the data from the 1710-1780 period onwards is included.) Whereas in the earliest period only one out of five *wh*-clauses was unmodified, the rate of unmodified pseudoclefts in written discourse today has risen to about three in five cases, or 60%.

Figure 3. Proportion of unmodified and modified *wh*-clauses
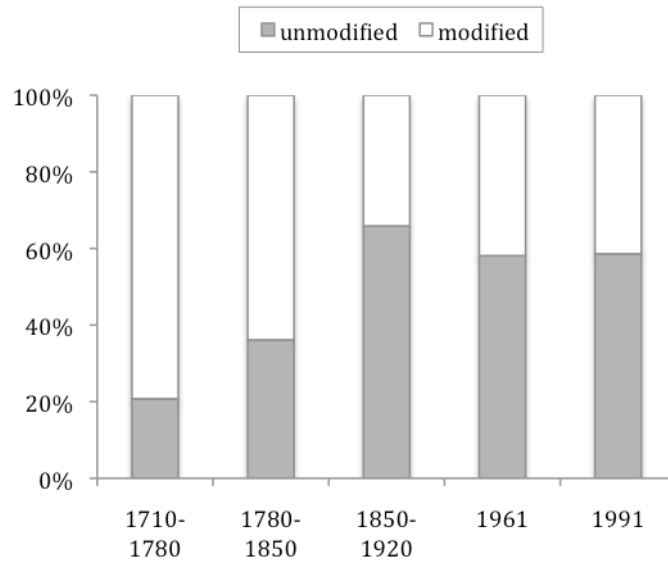
2.4 Wh-clause predicates

A second characteristic feature of early English pseudoclefts was that the types of predicates typically used in the *wh*-clause were chosen from a narrow semantic range. Table 2 lists the most frequent predicates in the 1710 - 1780 period. It shows that adjectives predominated. The frequently used verbs expressed similar evaluative notions as the adjectives. In general, the *wh*-clause expressed an evaluation of the focused constituent on a small number of semantic scales: quantity of some stated or understood property (*what is more, what will suffice*), exceptionality (*what was remarkable, what struck me*), pleasantness (*what I liked, what was worse*), and importance (*what matters, what is of importance*).

Table 2: Early *wh*-clause predicates (occurring two or more times)

1710 – 1780	
be more	10
surprise (me)	4
be remarkable	3
be worse	3
like	2
add to	2
be a hardship	2
be astonishing	2
be of importance	2
be ridiculous	2
be surprising	2
contribute	2
suffice	2

The early dominance of evaluative predicates provides another indicator of the construction's subsequent generalization. Tables 3, 4, and 5 show how the spectrum of *wh*-clause predicates expanded over time to include predicates that are less evaluative and more eventive. Utterance verbs (*say, mean, tell*) soon appear in the uppermost ranks, next to perception verbs (*see, hear*) and cognition verbs (*want, know*). Finally, the semantically most general verbs *do* and *happen* (shown in small capitals) enter the top frequency range, with *do* becoming the second most frequent predicate in the 20th century.

Tables 3, 4 and 5: Later *wh*-clause predicates

1780-1850		1850-1920		1961, 1991	
be worse	7	*want*	22	*be more*	18
be more	6	*be more*	19	DO	11
like	4	*say*	17	*need*	13
DO	4	DO	16	*say*	8
mean	4	*mean*	8	*mean*	7
want	4	*know*	8	*have (got)*	7
strike (me)	3	*see*	8	*matter*	7
tell	3	*be worse*	5	*see*	6
		like	4	*HAPPEN*	5
		propose	4		
		puzzle (me)	4		

The verbs *do* and *happen* are of particular interest in connection with pseudoclefts because in contemporary spoken English they alone make up the lion's share of all instances in usage (Collins 1991, Hopper 2004). Our own spoken English data also shows this strong skew towards *do* and *happen*, as can be seen in Table 6. It is interesting, then, that historically, and in terms of frequency, *do* and *happen* are relative latecomers.[2] Moreover, there seems to have been a stepwise introduction of them into the construction, with *do* preceding *happen*, as shown in Table 7. We will return to the relevance of this sequence in Section 4.

Table 6. Most frequent *wh*-clause predicates in present-day spoken American English (*Santa Barbara* corpus)

1980s (spoken)	
DO	55
HAPPEN	17
say	6
BE	5
be funny	3
be good	2
get [me]	2
remember	2
talk about	2

---

[2] Traugott (2008) discusses early examples of pseudoclefts with *do* in dramatic dialogue.

Table 7: Proportion of *do* and *happen* as *wh*-clause predicates in the written corpora

	1640-1710	1710-1780	1780-1850	1850-1920	1961, 1991
do	0	0	4	16	13
happen	0	0	0	2	6
Other predicates	3	106	115	231	247
Proportion *do/happen*	0%	0%	3%	7%	7%

The specialization of the construction to be used with *do* and *happen*, especially in spoken discourse, seems to be the final stage in the development of a mature pseudocleft.[3] A sign of even further generalization is the use in present-day spoken (but not written) English of the maximally general verb *be* in pseudoclefts, as in (12). The utterance was produced in reply to the question what a brand inspection is.

(12)   What it is is your only legal bill of sale, for horse or cattle, in the State of Colorado.

(SBCSAE)

To summarize the historical analysis, *wh*-clefts in English developed their modern usage potential gradually. The emergence of today's fully productive construction can be traced on the basis of several parameters, which provide measures of its degree of grammaticalization. First, the construction's status is reflected in the degree to which the relation between the *wh*-clause proposition and the preceding discourse was made explicit by the use of appropriate modifying expressions. Second, the types of predicates occurring in the *wh*-clause show that the construction could contain an increasing number of different open propositions. Originally, these were largely restricted to evaluations of a referent or state of affairs (e.g., *what surprised me…*). Later, predicates appear which express an event (e.g., *what happened was…*).

3.   Pseudoclefts in present-day English, Swedish and German

Before we discuss the motivations underlying the diachronic course of events, we take a look at pseudoclefts in two other languages, German and Swedish, in a synchronic, corpus-based comparison with English. The purpose of this analysis is to determine whether the parameters we used to quantify and chart the grammaticalization of pseudoclefts in Section 2 reflect idiosyncrasies of the history of English or whether they are more generally applicable. Can they be used to assess the status of differently developed pseudoclefts in other languages? If they are indeed more generally valid, perhaps reflecting aspects of a universal grammaticalization path, less developed pseudoclefts in other languages should echo the properties seen in English at earlier historical stages.

---

[3] The strong skew towards 'do' and 'happen' is also found in spoken Tagalog, where, judging by their discourse frequeny, pseudoclefts are at least as advanced as in English (Naonori Nagaya, p.c.).

It is well known that what can be expressed in form of a cleft in one language often is not, or not as easily, rendered as a cleft in another. The non-equivalence of even structural counterparts in closely related languages has been demonstrated in contrastive translation corpus studies. For example, M. Johansson (2001, 2002) examined the frequency with which clefts occurring in English novels are translated as clefts in the novels' Swedish editions. The rate at which they are rendered as clefts is generally low (see also S. Johansson 2001 for English, German and Norwegian). But while it is clear that not all clefts of the same general type, for example pseudoclefts, are grammatically equal, there has so far been no general grammaticalization-oriented attempt to work out what exactly their differences consist in and how to explain them. For example, in a recent cross-linguistic survey, Miller (2006) notes that while English has "the full range of clefts," other languages have only "a rudimentary cleft construction" (171). Miller's term "rudimentary" is intuitively appealing, but its implications are unclear. Some sort of development is implied but not made explicit. What exactly are such clefts lacking? And does their deficiency follow a general pattern?

The driving hypothesis behind our analysis is that a large amount of the cross-linguistic variability may be reduced to different degrees of grammaticalization. Many synchronic differences between the pseudoclefts of English, Swedish, and German, for example, might be the predictable effects of pseudoclefts at different points on a common developmental trajectory. Specifically, given their more rudimentary character, we predict that German and Swedish pseudocleft constructions should consistently score lower than the English pseudoclefts on our quantitative measures of grammaticalization. Also, to the extent that German and Swedish pseudoclefts themselves are at different developmental stages, our measures should enable us to rank them relative to each other.

Our choice of German and Swedish as points of comparison is of course also partially based on the availability of large amounts of usage data in the form of electronic corpora, which our methodology requires. We used four late 20th-century corpora, two each for Swedish and German, spoken and written discourse. The written data are taken from the *Stockholm Umeå Corpus* (Ejerhed et al. 2006) and the *HAMBURG* corpus (Hilpert 2004), which are exactly comparable to the LOB and FLOB corpora because they were constructed according to the same sampling design. The two spoken language corpora are the *Göteborg University spoken language corpus* (Allwood et al. 2000) and the *Freiburg* (aka. *Grundstrukturen*) corpus (Engel & Vogel 1975).[4]

Our operational definition of pseudoclefts had to be adjusted to the syntactic facts of German and Swedish, which are slightly different from those of English pseudoclefts. First, in both German and Swedish the presupposed part of the construction not only takes the form of a headless 'what'-relative (Ger. *was*, Swe. *vad*), but can also be a headed relative clause (Ger. *das was*, Swe. *det som* 'that which').

---

[4] Our German and Swedish spoken corpora are admittedly not as representative of casual conversation as, for example, the *Santa Barbara* corpus. They contain a substantial amount of academic discourse and other more formal speech events. However, given that pseudoclefts are known to be used more frequently in formal discourse, this should have increased their frequency, thus making the German and Swedish spoken data more similar to the English data. The large differences we find are therefore all the more remarkable.

(13)   German
<u>Das was</u> ich die ganze Zeit mit Ihnen erörtert habe war im Grunde...
'What we've been discussing all this time was basically...'

(Freiburg corpus)

(14)   Swedish
<u>Det som</u> händer är att det bubblar upp syra.
'What happens is that acid is bubbling up.'   (Göteborg corpus)

Swedish usage further includes the variants *det vad* ('that what'), *vad som* ('what which') and *det vad som* ('that what which'). The data clearly show that these variants are all used in a specificational function.[5]

Another point of difference is that German uses complex *wh*-words in relative clauses where the relativized argument is a prepositional object, e.g. <u>*worum es hier geht*</u> '<u>what</u> this is <u>about</u>' or <u>*wovon das abhängt*</u> '<u>what</u> this depends <u>on</u>'. To keep our analysis parallel to the English one, we included these complex *wh*-words if the relative clause containing them could be translated into English using *what* and a preposition.

Finally, in German as in Swedish the relative clause proposition may have a pronominal antecedent, leaving the relative clause, as it were, in a left-dislocated position.

(15)   German
Was uns bewegt, was uns Sorge macht, <u>das</u> ist eben dies, dass …
'What moves us, what worries us, (<u>that</u>) is just this, that …'

(Freiburg corpus)

(16)   Swedish
Det som egentligen är uppgiften <u>det</u> är att diskutera vad orden betyder.

'What really is the task (<u>that</u>) is to discuss what the words mean.'

(Göteborg corpus)

As in English, we included cases in which the relative clause and the focused constituent are not linked by a copula. We found them to be more common in German. An example is given in (17).

(17)   Was sehr viel wichtiger ist in unserem Zusammenhang, wenn wir …
'What's much more important in this connection, if we …'

(Freiburg corpus)

Methodologically, our corpus analysis followed the same, maximally inclusive search criteria as the historical analysis. We obtained complete concordances of the relevant *wh*-words, as well as Swedish *det som*, and then manually inspected each concordance line.

---

[5] This is not to say that such relatives may not be understood as predicational in other cases. In fact, the headed relatives show a slightly greater a tendency to be used in predicational constructions than the headless ones.

3.1 Discourse frequency

Figures 4 and 5 show that in terms of discourse frequency we find a consistent pattern GER < SWE < ENG on both frequency measures. Pseudoclefts are also consistently more common in spoken discourse.

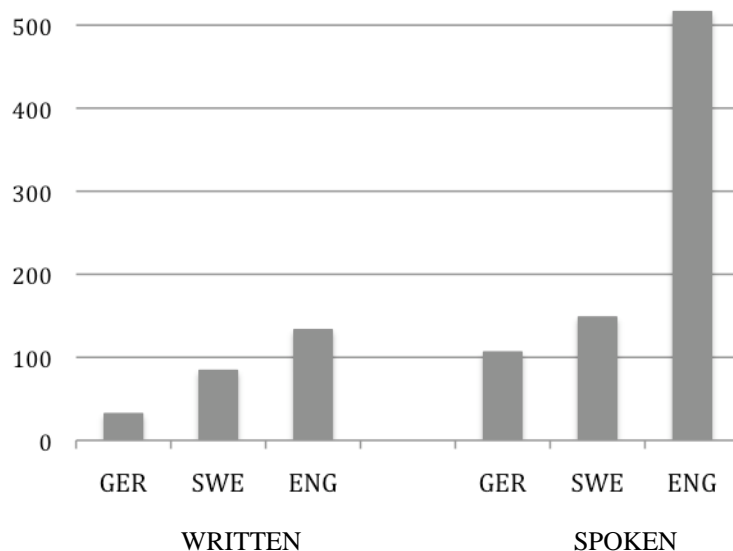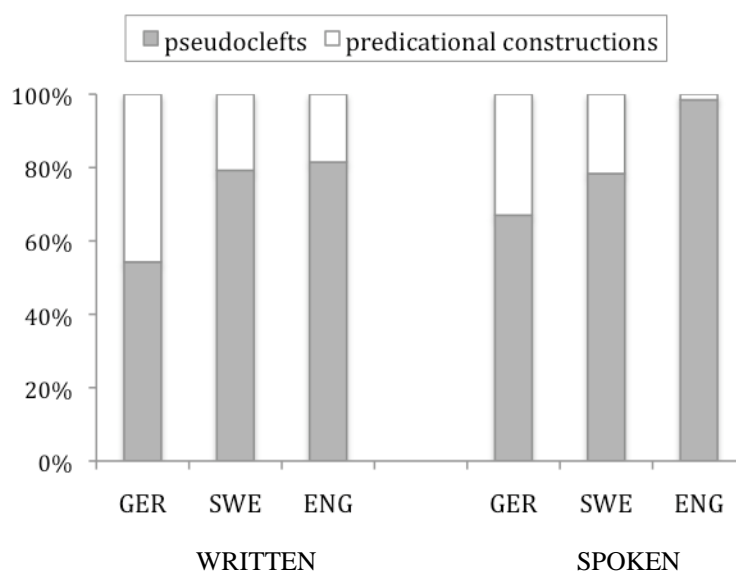Figure 4. Basic discourse frequency (number of occurrences per 1 million words)

Figure 5. Proportion of pseudoclefts and predicational copular constructions

3.2 Modification of the relative clause

Additional modification of the relative clause is very common in German, as seen for example in example (17) above. Table 8 shows the different types of modification, which by and large coincide with those found in English (cf. Table 1). Table 9 shows the modifiers found in the Swedish data, where modification is not quite as frequent.

Table 8. Types of modification of the relative clause in German

Comparison	*Was mir <u>viel mehr</u> auffiel ...*
Exceptional degree	*Was mich <u>besonders</u> gefreut hat ...*
Deixis	*Was <u>da</u> zu sehen ist ... Was wir <u>jetzt</u> brauchen ...*
Epistemic modality	*Was <u>sicher</u> nicht stimmt ...*
Anaphoric *so*	*Was mich <u>so</u> verwunderte ...*
Other anaphor	*Was ich <u>damit</u> meine ...*
Temporal adverbs	*Was man <u>dann</u> bekommt ...*
Addition	*Was ich <u>noch</u> sagen möchte ... Was <u>auch</u> fehlt ...*

Table 9. Types of modification of the relative clause in Swedish

Comparison	*Vad vi gör åt det <u>mer</u> ...*
Exceptional degree	*Det som går upp <u>reellt</u> ...*
Deixis	*Det som <u>sen</u> hände ...*
Epistemic modality	*Vad som <u>kan</u> hända ...*
Anaphoric *så*	*Det som är <u>så</u> synd ...*
Temporal adverbs	*Det som <u>fortfarande</u> skiljer det ...*
Addition	*Vad som syns <u>också</u> här ...*

As can be seen in Figure 6, the rate of unmodified pseudoclefts is consistently lowest in German. In spoken discourse we again find the pattern GER < SWE < ENG. Interestingly, there is no clear difference between the English and Swedish written data here. It appears that in Swedish writing unmodified relatives in pseudoclefts have already reached the ceiling level of about 60%, which we saw above for written English (cf. Figure 3).

Figure 6. Proportion of unmodified and modified relative clauses
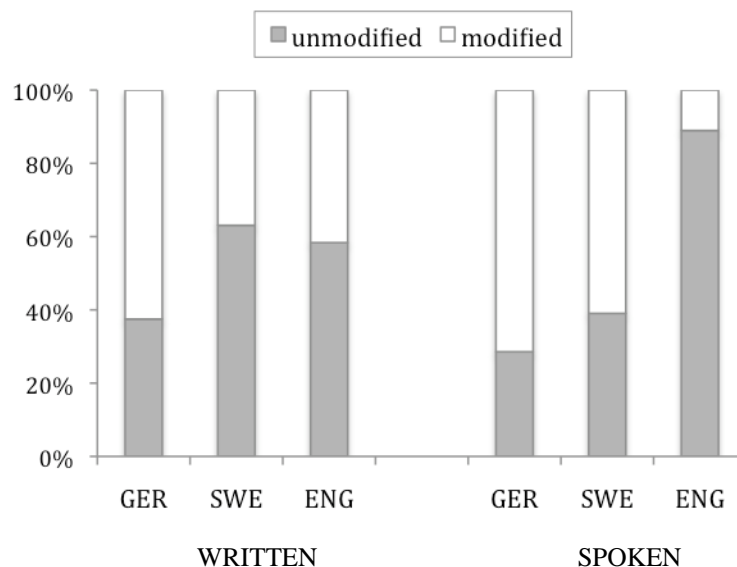
3.3 Wh-clause predicates

The predicates occurring in the relative clause provide more evidence for the same general pattern. Tables 10 and 11 list the most common predicates in German and their absolute frequency in the corpora. (The reason for the scarcity of the data is the low discourse frequency of pseudoclefts in German overall. Very few predicates occur more than once.) Note that evaluative predicates predominate ('be important', 'regret', 'puzzle, surprise') but that utterance and perception verbs ('say', 'mention', 'see') are well established. Conspicuously absent, at least among the high frequency predicates, are the equivalents of 'do' and 'happen'. These are almost completely absent in writing, and rare in spoken discourse (see also Weinert 1995; cf. Günthner 2007). In Swedish, by contrast, *göra* 'do' and *hända* 'happen' occupy the top frequency ranks in both the written and the spoken data, as shown in Tables 12 and 13. Other common predicates in Swedish include utterance verbs ('say', 'mean', and 'think' in the sense of expressing an opinion) and perception verbs ('see', 'distinguish').

Tables 10 and 11: Most frequent relative clause predicates in German

German written		German spoken	
wichtig sein 'be important'	2	*wichtig sein* 'be important'	5
erwähnen 'mention'	2	*sehen* 'see'	5
bleiben 'remain'	2	*bedauern* 'regret'	2
		verwundern 'puzzle, surprise'	2
		sagen 'say'	2

Tables 12 and 13: Most frequent relative clause predicates in Swedish

Swedish written		Swedish spoken	
göra 'do'	5	*göra* 'do'	20
handla om 'be concerned with'	5	*hända* 'happen'	15
hända 'happen'	4	*säga* 'say'	9
vara kvar 'be left, remain	3	*mena* 'mean'	8
utmärka 'distinguish'	3	*tycka* 'think'	6
		se 'see'	6

Figure 7 provides a quantitative comparison of the proportion of the verbs 'do' and 'happen' out of all attested predicates. It shows the by now familiar pattern GER < SWE < ENG, except that once more the difference between written Swedish and English is not as clear as expected.

Figure 7. Proportion of relative clause predicates 'do' and 'happen'
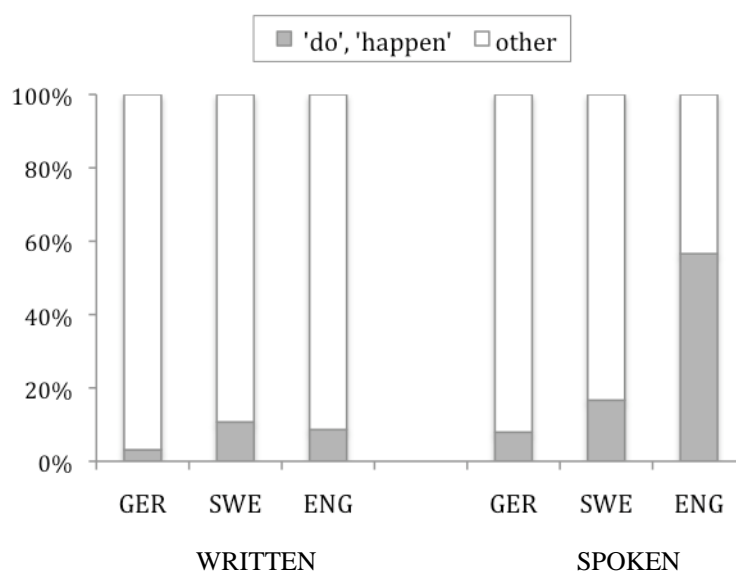
In summary, the quantitative and qualitative measures of grammaticalization that helped describe the diachronic data also capture relevant cross-linguistic differences, at least within the narrow range of Germanic languages considered here. The pseudoclefts of present-day German appear to be roughly at the stage of English pseudoclefts around the year 1800, while the Swedish pseudoclefts appear to be no more than one century behind the English ones. The striking resemblance of the historical and the cross-linguistic data suggests that the properties of pseudoclefts in individual languages can indeed to a large extent be regarded as cut-off points on a general grammaticalization continuum. The constraints which historically shaped the emergence of pseudoclefts in English are apparently the same as the constraints that "hold back" the corresponding constructions in German and Swedish today.

4.   Discussion

Our first question at the outset of this chapter was whether pseudoclefts arise instantaneously or gradually. The analysis has shown that they are the result of a gradual development that follows measurable stages. This finding does not necessarily invalidate Harris and Campbell's claim that cleft constructions are "easily added to grammars." After all, 300 years is not a very long time period in terms of diachronic syntax. And the reason for their fairly rapid growth may well lie in the availability of the syntactic components, as discussed in the Introduction. But the data show clearly that the pragmatic properties of a mature cleft construction do not automatically fall out in the process of combining a relative clause and a copular clause in a novel way. It takes a sustained process of generalization of the grammatical components, or rather, of their function as part of the emerging pseudocleft, for the construction to move beyond its more "exploratory" phase. In this restricted sense, pseudoclefts can therefore not be called "universally available."

More interesting than the rate of change is the particular course that the grammaticalization process takes, and what this reveals about the constraints the

emerging construction has to overcome on its way to becoming fully productive. Here we need to take a step back and reconsider what exactly it is that our measures of grammaticalization capture about pseudoclefts. Note that the changes we observed affect one particular component, the *wh*-clause. It starts out heavily constrained and then comes to be used more and more freely. In order to understand this process of generalization, we now briefly re-visit the information structure of cleft constructions, drawing primarily on Lambrecht's (1994, 2001) information structure framework.

4.1   Presuppositional structure and its pragmatic accommodation

As noted in Section 1, it is traditionally assumed that the use of cleft constructions requires interlocutors to share certain background assumptions. In pseudoclefts, the open proposition expressed in the *wh*-clause is *pragmatically presupposed*. Lambrecht (2001: 474) defines a *pragmatic presupposition* as

> [t]he set of propositions lexico-grammatically evoked in a sentence that the speaker assumes the hearer already knows or believes or is ready to take for granted at the time the sentence is uttered.

The set of pragmatic presuppositions conventionally associated with a grammatical construction constitute its *presuppositional structure*. To illustrate, the constructed exchange in (18) is a case in which the context clearly matches the pseudocleft's presuppositional structure.

(18)   A: Is your knee still giving you trouble?
        B: Actually, what I hurt is my ankle.

Speaker A's question establishes the right presupposition, viz. that speaker B was hurt.[6] Contexts that match a construction's presuppositional structure are referred to as appropriate *presuppositional situations*.

It is not difficult to imagine contexts that do not, or not necessarily, constitute appropriate presuppositional situations. One type, which has figured prominently in the synchronic pragmatic literature on clefts, is the use of them as "discourse openers." For example, Prince (1978) discusses the hypothetical utterances in (19), a professor's opening words in a lecture or course.

(19)   a.   *What one of my colleagues said this morning was …
        b.   What we're going to look at today (this term) is …

(Prince 1978: 889)

The lecture context rules out (19a) because the question of what a colleague may have said is not sufficiently predictable, or in Prince's words, "cooperatively assumable" (1978: 889). Yet (19b) is appropriate. In fact, as Prince points out, it would be appropriate in this context even if the listeners were not actually thinking about the contents of the class at all at the moment the sentence is uttered.

---

[6] More precisely, what is presupposed in (18) is not only that speaker B was hurt (Lambrecht's *K-presupposition*), but also that this knowledge has been activated in the minds of the interlocutors and is a predictable topic at the point of utterance (Lambrecht's *C-* and *T-presuppositions*; see Lambrecht 2001).

The possibility of using a cleft construction in the absence of a presupposition that it conventionally requires, as in (19b), is captured by a caveat in Lambrecht's definition of pragmatic presupposition (cited above). It states that hearers need to be at least "ready to take for granted" the relevant proposition. Drawing on Lewis (1979), Lambrecht refers to the act of going along with a speaker and acting as though a proposition were indeed shared as the *pragmatic accommodation* of presuppositional structure. In (19b), the pseudocleft's presuppositional structure can be accommodated easily. The open proposition 'today we're going to look at x' is so easy to "find" in the situational context that listeners can be expected to supply it themselves.

The option of relying on a certain amount of pragmatic accommodation means that the range of appropriate presuppositional situations for a particular cleft construction is not fixed but somewhat negotiable. Whether an open proposition can be *presented as presupposed* depends on the amount of accommodation listeners are able or willing to perform, given the nature of the proposition and the nature of the available linguistic and non-linguistic contextual clues. We believe that this inherent flexibility is the key to understanding the course of events in the history of the English pseudoclefts.

4.2   Modification of the wh-clause

Consider one more set of examples from the synchronic pragmatic literature on clefts. Declerck (1988) presents the invented utterance in (20) in support of his argument against Prince (1978) that pseudoclefts may even be used where the relevant presupposition is not in any way "cooperatively assumable." The intended context of (20) is the beginning of a speech. Later in his book, he provides another example, shown in (21), to illustrate the same point.

(20)   What I have <u>often</u> asked myself is how other linguists manage to keep abreast with the rapid developments in the different fields of linguistics while still finding time to go on writing articles themselves.

(21)   A: I hear you've got a job at Johnson's. A nice place that is. I suppose you're happy now?
        B: Well, I don't know. What I'd <u>really</u> like to do is run a business of my own.

(Declerck 1988: 213, 216, underlining added)

While the situational context may be of less help in these cases than in (19b), the principle of pragmatic accommodation still accounts for these cases. Specifically, what helps in accommodating the open propositions in (20) and (21) are the additional modifiers placed in the *wh*-clauses (whose role is not discussed by Declerck). To test their import, we invite the reader to judge the felicity of (20) and (21) if *often* and *really* were omitted. *Often* in (20) boosts the perceived relevance of the fact that the speaker has asked herself the particular question. Listeners are more likely to accept the presupposition 'I have asked myself x' out of the blue if it is a question that she has *often* asked herself. *Really* in (21) allows the proposition 'B would like to do x' to be accommodated more easily by indexing alternative values for the variable in the *wh*-clause, i.e. other career choices. One candidate is available in the linguistic context, viz. B's job at Johnson's. Thus, *really* facilitates the *ad hoc* construction of the required presupposition by providing a link to the prior discourse.

Modifying expressions like *often* and *really* are of course among the types of modification historically found in pseudoclefts (cf. Table 1). As discussed in Section 2.3, especially the early pseudoclefts often included such modifiers in the *wh*-clause to connect it to the ongoing discourse. We can now re-interpret the function of these modifiers as that of facilitating the pragmatic accommodation of the *wh*-clause proposition. Moreover, if the purpose of additionally modifying the *wh*-clause was indeed to facilitate the accommodation of the pseudocleft's presuppositional structure, the decreasing rate of occurrence of these modifiers over time suggests that their effect was most needed early on. This was followed by a period in which listeners could be increasingly counted on to accommodate the presuppositional structure of pseudoclefts even in their unmodified form. We will return to this point below.

4.3.   Wh-clause predicates

The gradual widening of the spectrum of attested *wh*-clause predicates discussed in Section 2.4 demonstrates the construction's capacity to express an increasing number of different open propositions. In the beginning, *wh*-clauses with evaluative predicates (e.g., *what's true…, what surprised me…*) were strongly preferred. In Kim's (1992, 1995) terminology, such *wh*-clauses are expressions of the speaker's *epistemic* and *affective stance*. Interpreted in terms of the notion of pragmatic accommodation, this finding suggests that such predicates, or rather the propositions expressed by the *wh*-clauses containing them, were easiest to accommodate pragmatically. Why should pragmatic accommodation be more readily secured in these cases? We suggest that this is because evaluations are grounded in the speaker's subjective beliefs, attitudes, and emotional reactions. Unlike events in the world, evaluative reactions can be taken for granted in many more situations.

It is probably no coincidence that Kim (1992) points precisely to an example involving an affective stance predicate in his critique of the relevance of traditional information structure categories, like presupposition and focus, for the analysis of pseudoclefts in conversation. Regarding the pseudocleft in (22), which occurred in a conversation, he points out that "the information contained in the initial WH-clause can in no way be related to some preceding context" (1992: 23).

(22)   The pay goes up- <u>what I'm surprise of</u> is the pay goes up to sump'n like two fifty an hour.

While we agree that the talk preceding (22) (not fully cited here) indeed does not include any prior indication of the speaker's surprise, we would argue that in a case like this the relevant open proposition is simply very easy to accommodate, so much so that no linguistic context is needed. The example thereby demonstrates the advantage of evaluative predicates.

Another reason for maintaining that some degree of accommodation is always involved is that the later, non-evaluative predicates come to be used in the construction in a particular historical sequence, which would otherwise remain unexplained (cf. Section 2.4). We can see this particularly clearly in the case of the verbs *do*, *happen*, and *be* The diachronic order in which they become available in pseudoclefts illustrates the general observation that the *wh*-clause proposition is initially more contextually dependent, i.e. that it needs to echo established aspects of the discourse to a greater extent. Note that the three verbs enter the construction in an order that reflects their degree of semantic schematicity.

$$(23) \quad \text{i.} \quad \text{`do'} \quad \quad \text{dynamic event, participants specified}$$

(23) i.   'do'   dynamic event, participants specified
      ii.   'happen' dynamic event, participants unspecified
      iii. 'be'   any event or state of affairs

Although *do* is already semantically highly general, in order of magnitude it is the most lexically rich of the group. It covers only dynamic events and requires the participants in that event to be made explicit (e.g., *what we did was...*). The verb *happen*, by contrast, requires no mention of the participants (e.g., *what happened was...*). Finally, the verb *be* is maximally general and not restricted to any particular event type (e.g., *what is was was (that)...*).[7]

In summary, the extension of the construction to additional *wh*-clause predicates proceeded in such a way that the range of open propositions expressed in the *wh*-clause came to include cases which were increasingly less easy to accommodate pragmatically. As seen with the additional modifiers placed in the *wh*-clause, this suggests that listeners could be increasingly relied on to routinely perform this task.

5.   Conclusion: The conventionalization of pragmatic accommodation

The general picture emerging from the above discussion is that from the beginning of their existence an important function of pseudoclefts was to *deliberately introduce* presuppositions into the discourse, i.e. to *create* desired presuppositional situations. The construction's presuppositional structure was regularly stretched beyond its limits by relying on the addressee's capacity to accommodate increasingly less clearly established, but still recoverable presuppositions. The long-term effect that such usage can have on grammatical constructions has been noted by Lambrecht, although not in connection with pseudoclefts.

> The pragmatic accommodation of certain presuppositional structures may to a greater or lesser extent become CONVENTIONALIZED and eventually GRAMMATICALIZED ... It can happen that the presuppositional structure of a frequently used construction is exploited so regularly that it loses some of its force, sometimes resulting in a new meaning for the construction (1994: 70).

Thus, in conclusion, we propose that the degree to which a pseudocleft's presuppositional structure is conventionally accommodated pragmatically by listeners is the key parameter along which these constructions vary historically. A more grammaticalized pseudocleft is treated by listeners, as it were, more leniently. Listeners are ready to take for granted open propositions that they would not accept in the case of a more rudimentary pseudocleft. It remains to be worked out how it is possible for there to be construction-specific degrees to which pragmatic accommodation occurs. We hope to address this issue in future research.

---

[7] The fact that pseudoclefts are compatible with *be* is particularly interesting because in many of these instances the pseudocleft combines with an "inferential" cleft, i.e. a constructions of the form *it's that*. Inferential clefts operate on antecedents in discourse which are typically very low in accessibility, viz. inferences (see Koops 2007). That inferential clefts nevertheless combine with pseudoclefts bears witness to the high degree of generalization of the English pseudocleft construction.

References

Ball, C. 1991. The Historical Development of the *it*-cleft. PhD dissertation, University of Pennsylvania.

Ball, C. 1994. The origins of the informative-presupposition *it*-cleft. *Journal of Pragmatics* 22: 603-628.

Collins, P. 1991. *Cleft and Pseudo-cleft Constructions in English.* London, New York: Routledge.

Declerck, R. 1988. *Studies in Copular Sentences, Clefts, and Pseudo-clefts.* Dordrecht: Foris.

Givón, T. 1979. *On Understanding Grammar.* New York: Academic Press.

Günthner, S. 2007. *Was ihn trieb, war vor allem Wanderlust*: Pseudocleft-Konstruktionen im Deutschen. In *Konstruktionen in der Interaktion*, S. Günthner & W. Imo (eds), 59-90. Berlin: Mouton de Gruyter.

Harris, A. & Campbell, L. 1995. *Historical Syntax in Cross-linguistic Perspective.* Cambridge: Cambridge University Press.

Heine, B. & Reh, M. 1984. *Grammaticalization and Reanalysis in African Languages.* Hamburg: Buske.

Hopper, P. J. 2001. Grammatical constructions and their discourse origins: Prototype or family resemblance? In *Applied Cognitive Linguistics*, M. Pütz, S. Niemeier & R. Dirven (eds), 109-129. Berlin: Mouton de Gruyter.

Hopper, P. J. 2004. The Openness of Grammatical Constructions. *Papers from the 40th Meeting of the Chicago Linguistic Society*.

Johansson, M. 2001. Clefts in contrast: a contrastive study of *it*-clefts and *wh*-clefts in English and Swedish texts and translations. *Linguistics* 39: 547-582.

Johansson, M. 2002. Clefts in English and Swedish. A Contrastive Study of IT-clefts and WH-clefts in Original Texts and Translations. PhD dissertation, Lund University.

Jones, L. B. & Jones, L. K. 1985. Discourse functions of five English sentence types. *Word* 36: 1-21.

Kim, K.-H. 1992. WH-Clefts and Left-Dislocation in English Conversation with Reference to Topicality in Korean. PhD dissertation, University of California, Los Angeles.

Kim, K.-H. 1995. Wh-clefts and left-dislocation in English conversation. In *Word Order in Discourse*, P. Downing & M. Noonan (eds), 247-296. Amsterdam: John Benjamins.

Koops, C. 2007. Constraints on inferential constructions. In *Aspects of Meaning Construction,* G. Radden et al. (eds), 207-224. Amsterdam: John Benjamins.

Koops, C. & Ross-Hagebaum, S. 2008. Information structure and discourse function of amalgam *wh*-clefts. *Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society*.

Lambrecht, K. 1994. *Information Structure and Sentence Form.* Cambridge: Cambridge University Press.

Lambrecht, K. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39: 463-516.

Lewis, David. 1979. Scorekeeping in a language game. In *Semantics from Different Points of View*, R. Bäuerle, U. Egli & A. von Stechow (eds), 172-187. Berlin, Heidelberg, New York: Springer Verlag

Miller, J. 2006. Focus in the languages of Europe. In *Pragmatic Organization of Discourse in the Languages of Europe*, G. Bernini & M. Schwartz (eds), 121-214. Berlin: Mouton de Gruyter.

Prince, E. 1978. A comparison of wh-clefts and it-clefts in discourse. *Language* 54: 883-906.

Traugott, E. C. 2008. *All that he endeavoured to prove was...*: On the emergence of grammatical constructions in dialogual and dialogic contexts. In *Language in Flux: Variation, Change and Evolution,* R. Kempson & R. Cooper (eds). London: College Publications.

Weinert, R. 1995. Focusing constructions in spoken language. Clefts, Y-movement, Thematization and Deixis in English and German. *Linguistische Berichte* 159: 341-369.

Weinert, R. & Miller, J. 1996. Cleft constructions in spoken language. *Journal of Pragmatics* 25: 173-206.

Corpora used

Allwood J., Björnberg, M., Grönqvist, L., Ahlsén, E. & Ottesjö, C. 2000. The Spoken Language Corpus at the Linguistics Department, Göteborg University. *Forum Qualitative Social Research* 1/3.

Berens, F. J., Jäger, K.-H., Schank, G. & Schwitalla, J. 1976. *Projekt Dialogstrukturen. Ein Arbeitsbericht*. München: Hueber.

De Smet, H. 2005. A corpus of Late Modern English. *ICAME Journal* 29: 69-82.

Du Bois, J., Chafe, W., Meyer, C. & Thompson, S. 2000. *Santa Barbara Corpus of Spoken American English*. Part 1. Philadelphia, PA: Linguistic Data Consortium.

Du Bois, J., Chafe, W., Meyer, C., Thompson, S. & Martey, N. 2003. *Santa Barbara Corpus of Spoken American English*. Part 2. Philadelphia, PA: Linguistic Data Consortium.

Du Bois, J. & Englebretson, R. 2004. *Santa Barbara Corpus of Spoken American English*. Part 3. Philadelphia, PA: Linguistic Data Consortium.

Du Bois, John & Englebretson, R. 2005. *Santa Barbara Corpus of Spoken American English*. Part 4. Philadelphia, PA: Linguistic Data Consortium.

Ejerhed, E., Källgren, G. & Brodda, B. 2006. *Stockholm Umeå Corpus, Version 2.0.* Stockholm: University of Stockholm, Department of Linguistics.

Engel, U. & Vogel, I. (ed.) 1975. *Gesprochene Sprache. Bericht der Forschungsstelle Freiburg*. [Forschungsberichte des Instituts für deutsche Sprache 7]. Tübingen: Narr.

Hilpert, M. 2004. *The HAMBURG corpus*. Houston, TX: Rice University.

Hofland, K., Lindebjerg, A. & Thunestvedt, J. 1999. *ICAME collection of English language corpora.* 2nd edition. Bergen, Norway: HIT Centre, University of Bergen.

Kroch, A., Santorini, B. & Delfs, L. 2004. *Penn-Helsinki Parsed Corpus of Early Modern English*. http://www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-1/