

1. Introduction

This article concerns the variation and change in long-distance (LD) dependencies in Dutch. Both historical and contemporary corpus data dealing with these constructions is discussed. Recently, it has been argued that LD-movement constructions are formed based on a fixed formula (cf. Dąbrowska 2004, 2008; Verhagen 2005, 2006). The reason for assuming this is that corpus data demonstrate that LD wh-questions show very limited variation in the domain of the matrix clause, and hence seem to conform to a general template. In this paper, we present a number of counterarguments against such an analysis.

The outline of this article is as follows. First, the four types of LD-movement constructions that are central to this paper are treated. Next, the data discussed in Dąbrowska (2004, 2008) and Verhagen (2005, 2006) is presented as well as some of the main claims these authors put forward. Subsequently, we present our own corpus data, which we argue forms evidence against the analogy analysis of LD-movement constructions. We point out some of the factors we believe cause the limited variation in the matrix clause of LD wh-questions, and round off with a general conclusion.

2. LD-movement constructions

LD-movement has been at the heart of generative grammar over the past few decades. Traditionally, four types of constructions are considered to involve this kind of A'-movement, namely wh-questions, relatives, topicalization constructions and comparatives (cf. Chomsky, 1977). These constructions are illustrated in (1) – (4), respectively

- (1) *Wh-movement*
[_{CP} Who do you think [_{CP} John will kiss t_{who}]]
- (2) *Relativization*
[_{CP} That is the girl who I think [_{CP} John will kiss t_{who}]]
- (3) *Topicalization*
[_{CP} The girl I think [_{CP} John will kiss t_{the girl}]]
- (4) *Comparatives*
[_{CP} John has kissed more girls [_{CP} than OP I think Peter did t_{OP}]]

Especially within generative frameworks, LD-movement is seen as a productive rule in which an element is moved from a subordinate clause into a higher clause. For example, in (1), the wh-phrase *who*, which is the object of the subordinate verb, has moved to the left periphery of the matrix clause. The reason for treating these constructions as one and the same is that they behave alike in many respects. In all cases, movement leaves behind a gap, and proceeds in intermediate steps. Furthermore, all four constructions are sensitive to the same kind of interveners.

¹ Corresponding author: Ankelien Schippers (a.schippers@rug.nl)

3. *The analogy account*

Contrary to what is assumed within the generative framework, it has recently been argued that LD-movement (specifically LD wh-movement) does not involve a productive rule at all, but rather that these constructions are based on a general template (cf. Dąbrowska, 2004, 2008; Verhagen, 2005, 2006). This analysis will be referred to as the analogy account. The idea is that any LD-construction departing from the general template is created by analogy to this template. This hypothesis springs from the observation that naturally occurring examples of LD wh-questions show very little variation regarding their type of matrix predicate and subject. Dąbrowska and Verhagen report that in English, the construction is almost exclusively attested with the matrix verb *think* or *say*, the auxiliary *do* and a 2nd person pronoun as the matrix subject. Dąbrowska (2004) looked at the Manchester corpus and found that 96% of the LD wh-questions had the matrix verb ‘think’ or ‘say’. Furthermore, 91% of the occurrences had ‘you’ as the subject and 99 % had some form of ‘do’ in the auxiliary position. Dąbrowska (2004) further looked at the CHILDES-data and found that 47 out of 49 occurrences of long-distance wh-questions were of the form “WH do you think S?” (where ‘S’ stands for ‘subordinate clause’). In Dąbrowska (2008), additional data from the British National Corpus (BNC) is discussed. She reports that 70 % of the LD wh-questions in the spoken part of the BNC also have the form “WH do you think S?”. Similar findings are reported in Verhagen (2005) and (2006) for the Brown corpus: out of 11 occurrences, 10 had the matrix verb ‘think’ and 1 ‘say’; 9 had the matrix subject *you*, and 10 occurred with a form of *do* as the auxiliary. In Verhagen (2005) and (2006), it is furthermore pointed out that Dutch shows a similar pattern. Verhagen searched the digital version of the newspaper *De Volkskrant* and the Eindhoven corpus for LD wh-questions. In the Eindhoven corpus, 6 out of 6 occurrences showed up with the matrix verb *denken* ‘think’ and a 2nd person personal pronoun. Data from the *Volkskrant* showed that 34 out of 43 occurrences had the matrix verb *denken* ‘think’, 5 *willen* ‘want’ and 4 *zeggen* ‘say’ or *vinden* ‘find’. Furthermore, 36 occurrences had a 2nd person personal pronoun as the matrix subject.

Based on these observations, Dąbrowska (2004, 2008) and Verhagen (2005, 2006) argue that LD-movement constructions are stored as fixed formulas as in (5a) below for English and (5b) for Dutch, and are created by analogy to this formula.

(5a) [WH do you think/say [S ...]]

(5b) [WH *denk je [dat ...]*
WH think you that

On this account, the more an LD-movement construction departs from this formula, the less acceptable it will be. The limited variation in LD wh-questions indeed suggests that the construction is not as productive as a purely formal account would predict. However, as is pointed out in the next sections, our data shows considerable variation in the type of matrix predicates and subjects that may show up in LD-movement constructions.

4. *Dutch diachronic corpus data*

The corpus data presented in the studies by Dąbrowska (2004, 2008) and Verhagen (2005, 2006) are very limited in that they capitalize on one type of LD-movement, namely in wh-questions. However, our data contain occurrences of four types of LD-movement constructions. In total, the data consist of 1734 occurrences, of which the oldest example is from the beginning of the 14th century. The data were collected by the second author by

manually inspecting texts for LD-movement constructions.² Table (1) shows the number of occurrences for each type of movement.

Table 1: Occurrences per type of construction

Type of construction	Frequency
Wh-questions	562
Relatives	872
Topicalization	196
Comparatives	104
Total	1734

We first discuss the type of matrix predicates, and then focus on the type of matrix subjects that are attested in the data.

4.1 Matrix predicates

Table (2) shows the frequencies and relative frequencies for the 20 most frequent verbs. The data show a wide variety of predicates: in total, 143 different matrix predicates were attested. In accordance with Verhagen’s finding, our data also show that the matrix verbs *denken* and *willen* are most frequent. However, it turns out that they are specifically frequent in wh-questions, and not so much in the other types of constructions. For example, in relative constructions, *zeggen* and *weten* are most frequent. This latter verb virtually does not show up in LD wh-questions. The reason for this is that *weten* is a factive verb, and when used to question over something, must combine with an interrogative complement. LD wh-questions, however, may only have a non-interrogative complement clause. This suggests the choice of matrix verb is partially influenced by semantic/pragmatic factors.

Furthermore, the data show that while wh-questions and comparatives indeed show a limited variety of matrix verbs, relatives and topicalization constructions surface with a wider variety of matrix verbs. Importantly, it appears that the fact that LD wh-questions show such limited variation regarding their type of matrix predicate is not a feature of LD-movement constructions in general, but rather of specific types of LD-movement (e.g. wh-questions). This speaks against the analogy account of Dąbrowska and Verhagen, which predicts all LD-movement constructions should show limited variation regarding their matrix predicate and subject.

Further evidence against the analogy account is presented in Ambridge and Goldberg (2008). As pointed out in the introduction, one of the predictions the analogy account makes is that the more an LD-movement construction departs from the general template, the less acceptable it will be. Ambridge and Goldberg tested this claim by collecting acceptability judgments on LD wh-questions. Their study showed that the acceptability of the constructions strongly correlated with the degree of backgroundedness of the complement clause, and not so much with whether the constructions were similar to a general template. Importantly, Ambridge & Goldberg did not find any significant correlation between the acceptability of a LD wh-question and the semantic similarity of its matrix verb to frequent verbs like *think* and *say*. Hence, their results speak against the analogy account.

² For more details, see Hoeksema & Schippers (2009) and Schippers (2009).

Table 2: Matrix verbs per type of construction*

Predicate	Wh	% of Wh	Rel	% of Rel	Top	% of Top	Com	% of Com	Total	% of Total
<i>denken</i> think	325	57,8	102	11,7	21	10,7	26	25,0	474	27,3
<i>willen</i> want	119	21,2	41	4,7	9	4,6	10	9,6	179	10,3
<i>zeggen</i> say	35	6,2	89	10,2	22	11,2	6	5,8	152	8,8
<i>weten</i> know	3	0,5	106	12,2	15	7,7	3	2,9	127	7,3
<i>menen</i> think	21	3,7	77	8,8	13	6,6	5	4,8	116	6,7
<i>hopen</i> hope	7	1,2	52	6,0	6	3,1	7	6,7	72	4,2
<i>zien</i> see	1	0,2	35	4,0	7	3,6	2	1,9	45	2,6
<i>geloven</i> believe	0	0,0	25	2,9	12	6,1	4	3,8	41	2,4
<i>vinden</i> consider	18	3,2	19	2,2	5	2,6	5	4,8	47	2,7
<i>wensen</i> wish	3	0,5	23	2,6	8	4,1	1	1,0	35	2,0
<i>vrezen</i> fear	0	0,0	15	1,7	10	5,1	2	1,9	27	1,6
<i>oordelen</i> judge	1	0,2	15	1,7	3	1,5	1	1,0	20	1,2
<i>begrijpen</i> comprehend	0	0,0	12	1,4	6	3,1	2	1,9	20	1,2
<i>vermoeden</i> suspect	1	0,2	14	1,6	3	1,5	1	1,0	19	1,1
<i>horen</i> hear	0	0,0	10	1,1	2	1,0	2	1,9	14	0,8
<i>verzoeken</i> request	0	0,0	13	1,5	1	0,5	0	0,0	14	0,8
<i>verwachten</i> expect	9	1,6	8	0,9	1	0,5	1	1,0	19	1,1
<i>vertrouwen</i> trust	0	0,0	8	0,9	5	2,6	1	1,0	14	0,8
<i>verzekeren</i> ensure	0	0,0	5	0,6	6	3,1	2	1,9	13	0,7
<i>beweren</i> claim	0	0,0	7	0,8	0	0,0	3	2,9	10	0,6
other types of predicates (119)	19	3,4	196	22,5	41	20,9	20	19,2	276	15,9
Total	562	100	872	100	196	100	104	100	1734	100

* Abbreviations: Wh = Wh-questions, Rel = Relative, Top = Topicalization, Com = Comparative.

4.2 Type of matrix subject

Further evidence against the analogy account comes from the fact that the four types of LD-movement also occur with a variety of matrix subjects. Table (3) shows for each construction the type of matrix subject.

Table 3: Type of matrix subject per construction

Matrix subject	Type of construction				Total
	Wh-questions	Relatives	Topicalization	Comparatives	
1SG pronoun	23	299	111	38	471
1PL pronoun	4	48	15	3	70
2SG pronoun	445	52	4	8	509
2PL pronoun	13	3	0	0	16
3SG pronoun	36	165	15	17	233
3PL pronoun	7	60	6	5	78
INDEF pronoun	1	13	1	0	15
No overt subject	30	100	18	24	172
Full NP	3	132	26	9	170
Total	562	872	196	104	1734

As can be seen from Table (3), 2nd person personal pronouns are indeed most frequent. However, this is only due to the fact that they are so frequent for wh-questions. The other three constructions are much more frequent with 1st person and 3rd person personal pronouns.

The reason why wh-movement constructions mainly show up with 2nd person personal pronouns seems to be pragmatic in nature. First of all, personal pronouns are far more frequent than full noun phrases (cf. Howe, 1996). Second, most matrix predicates in wh-questions are mental verbs (verbs like ‘think’ and ‘hope’). From a pragmatic view, it is much more natural to ask a question about someone’s thoughts/hopes to an addressee, than to oneself or a third party. Furthermore, the reason why pronouns instead of full NPs are used in this case is most likely due to the fact that it is much more natural to refer to an addressee by means of a personal pronoun than by means of a full noun phrase (e.g. a proper name). Hence, the predominance of 2nd person personal pronouns appears to be due to pragmatic reasons only.

Further evidence that the matrix subject is not part of a fixed formula is put forward in Dąbrowska (2008). She collected acceptability judgments on LD wh-questions, where the focus was on whether the similarity of the matrix verb and subject to the general template had an influence on the acceptability. Whereas she did find a significant effect for the type of matrix verb, it appeared that participants did not make a difference between whether the matrix subject was a personal pronoun or a full NP. However, the analogy account would predict participants to have a preference for the most frequent pattern, hence for 2nd person personal pronouns. This is further evidence against the assumption that the predominance of 2nd person pronouns is due to the templatic nature of LD-movement constructions.

5. Diachronic development of LD-movement in Dutch

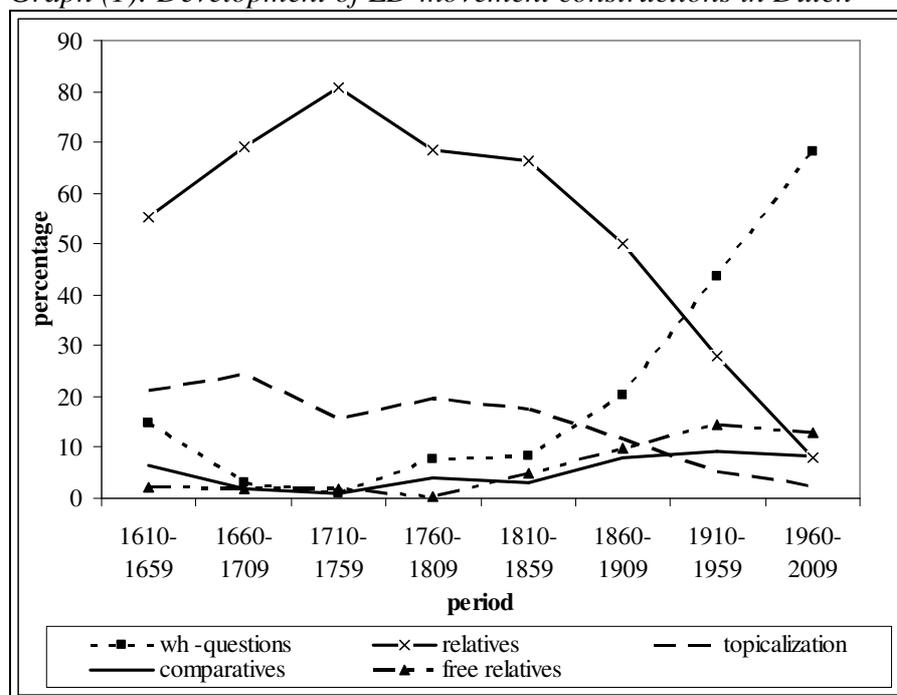
Although we showed previously that the limited variation in the matrix clause of LD-movement constructions is something which is not typical of these constructions in general, the question that remains is why LD wh-questions and comparatives only occur with such a limited variety of matrix predicates, contrary to relatives and topicalization constructions.

Interestingly, relatives (in particular headed relatives) and topicalization constructions appear to differ in another respect from wh-questions and comparatives as well. The historical data shows that headed relatives and topicalization constructions show a relative decrease in frequency compared to wh-questions, free relatives and comparatives, starting around the middle of the 19th century. This is shown in Table (4) and graph (1).³ Graph (1) shows the relative frequencies for each type of movement per period. These were computed by determining for each period the percentage of occurrences relative to the total number of LD-movement occurrences in that period. It is clear that especially LD wh-questions show a strong increase over the past two centuries, while topicalization constructions and headed relatives start to decrease.

Table 4: Frequency LD-movement constructions 1610 - present

Period	Type of construction					Total
	Wh-questions	Headed relatives	Topicalization	Comparatives	Free relatives	
1610 - 1659	7	26	10	3	1	47
1660 - 1709	5	114	40	3	3	165
1710 - 1759	1	88	17	1	2	109
1760 - 1809	17	154	44	9	1	225
1810 - 1859	14	111	29	5	8	167
1860 - 1909	45	111	26	18	22	222
1910 - 1959	403	47	14	49	77	590
1960 - present	559	694	188	102	136	1679

Graph (1): Development of LD-movement constructions in Dutch



³ Free relatives are treated separately from headed relatives for reasons that will become apparent in a moment.

It appears that the decrease in headed relatives and topicalization constructions is due to the replacement by an alternative construction, called resumptive prolepsis. This construction is discussed extensively in Salzmann (2006), and does not seem to involve LD-movement proper: it occurs with all sorts of islands, and the gap site is filled by a resumptive pronoun, which is otherwise not allowed in Dutch. An example is in (6):

- (6) [_{CP} *De manⁱ van wieⁱ ik denk* [_{CP} *dat hijⁱ de fiets gestolen heeft*]]
 the man of whom I think that he the bike stolen has
 ‘The man whom I think stole the bike’

As pointed out in Salzmann (2006), the resumptive prolepsis construction can be used as an alternative for LD-relativization, topicalization and wh-movement, but cannot be used as an alternative for LD-comparatives and free relatives. The reason for this is that resumptive prolepsis is only possible when the noun phrase is referential/d-linked. Furthermore, for reasons that are not entirely clear to us, resumptive prolepsis is most natural for topicalization and relativization, and much less for wh-questions. Hence, resumptive prolepsis is not normally used as an alternative for LD wh-movement. Since resumptive prolepsis is only a suitable alternative in the environment of relativization and topicalization, it explains why specifically these constructions decline in frequency.

While we do not have any conclusive evidence to prove that the resumptive prolepsis construction has replaced LD-movement constructions, there are several observations that suggest this is indeed the case. First of all, it appears that a similar process has taken place in German. It has been reported by a number of authors that LD-movement in German has started to decrease around the same time as in Dutch, namely around the middle of the 19th century. These authors have also pointed out that in German, LD-dependencies are instead formed by using alternatives, one of them being resumptive prolepsis (cf. Blatz, 1896; Paul, 1920; Behaghel, 1928; Ebert, 1973; Andersson & Kvam, 1984; Lühr, 1998 and Salzmann, 2006).

Further evidence that LD-movement constructions decrease due to the availability of alternatives is provided by the decline of LD wh-movement in German. German, contrary to Dutch, has alternatives to form LD wh-questions, such as partial wh-movement.⁴ Consequently, LD wh-movement also started to decline in German, whereas in Dutch, it can still be frequently attested. Furthermore, the idea that LD-movement constructions are replaced by alternative constructions is also corroborated by the fact that free relatives, contrary to headed relatives, do not show a decrease in frequency. This strongly suggests the decline of headed LD-relatives is not something inherent to LD-relativization, but rather directly related to the possibility of using an alternative construction.

In sum, the constructions that do not have a proper alternative (LD wh-questions, free relatives and comparatives), do not show a decrease in frequency. On the contrary, these constructions show a relative increase (cf. Schippers, 2009). This is particularly true for LD wh-questions, which show a very strong relative increase over the past few centuries. Interestingly, this is also precisely the construction that shows very limited variation in the domain of the matrix clause. Hence, while the relatively high frequency of LD wh-questions in Dutch suggests it is quite a productive construction, the limited variation in the domain of its matrix clause suggests otherwise.

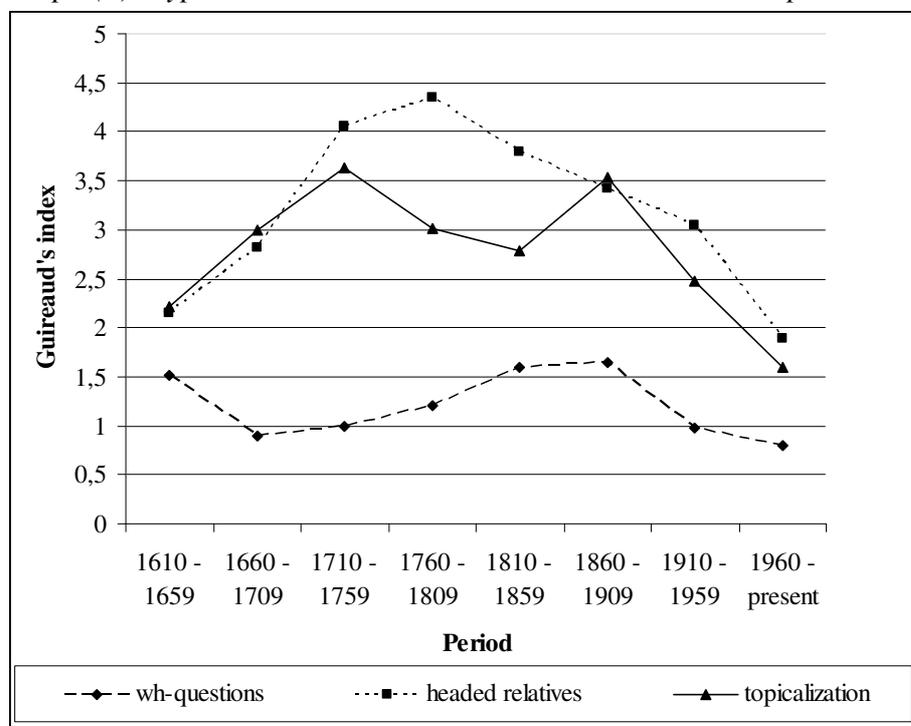
⁴ Partial wh-movement can be attested in Dutch as well, but has a marginal status and is normally not used as an alternative to LD wh-movement in the standard language (cf. Schippers, 2009)

There is some evidence that the limited variation in matrix predicates has a diachronic dimension as well. This becomes obvious when we look at the type/token ratios of the constructions under consideration. While type/tokes ratios are not a very reliable measure of variation, they do give a general idea of the degree of variation. Table (5) and Graph (3) give the type token ratios for wh-movement, headed relatives and topicalization constructions. Comparative constructions and free relatives were not taken into consideration, since there is just too little data per period to deduce anything meaningful from them. To adjust for the fact that the samples are not the same for each period and type of movement, Guiraud's index was used. This is the type/token ratio where the types are divided by the square root of the tokens. Table (5) and Graph (2) show that the type/token ratio for all three constructions declines, again around the middle of the 19th century. This is the same period at which relative and topicalization constructions also generally start to decline in frequency.

Table 5: Type/token ratio's per period

Period	wh-questions		headed relatives		topicalization	
	type/token	$\frac{type}{\sqrt{tokens}}$	type/token	$\frac{type}{\sqrt{tokens}}$	type/token	$\frac{type}{\sqrt{tokens}}$
1610 - 1659	4/7	1,51	11/26	2,16	7/10	2,21
1660 - 1709	2/5	0,89	30/114	2,81	19/40	3,00
1710 - 1759	1/1	1,00	38/88	4,05	15/17	3,64
1760 - 1809	5/17	1,21	54/154	4,35	20/44	3,02
1810 - 1859	6/14	1,60	40/111	3,80	15/29	2,79
1860 - 1909	11/45	1,64	36/111	3,42	18/26	3,53
1910 - 1959	8/67	0,98	20/43	3,05	7/8	2,47
1960 - present	16/403	0,80	13/47	1,90	6/14	1,60

Graph (2): Type/token ratio's wh-movement, relatives and topicalization



Taken together, the picture that emerges from this is that LD-movement in Dutch is generally becoming a less productive phenomenon. Relative and topicalization constructions are showing a decline that is most likely due to replacement by the resumptive prolepsis construction. Furthermore, this decline is mirrored by the decreasing variation in type of matrix predicates. Hence, notwithstanding the fact that LD wh-questions are increasing in frequency, the productivity of this construction actually appears to decrease. This is something which was already noted by Verhagen, and which can also be witnessed in our data: especially in more recent periods, the variety of matrix predicates in this construction is limited. However, this is not something that is particular to LD-movement itself, but rather to specific LD-movement constructions, such as wh-questions, and likely subject to diachronic change.

6. Conclusion

In this paper, the variation and change in Dutch LD-dependencies was discussed. We argued that the limited variation attested in LD wh-questions is not simply due to the fact that these constructions are based on a general template, since LD-movement constructions other than wh-questions show much more variation. Furthermore, we argued that there is evidence that LD-movement is less productive as the result of a diachronic process. We also pointed out some pragmatic and semantic issues that influence the type of matrix predicate and subject. All in all, our data provide evidence against an analogy account of these constructions, since they show a considerable amount of lexical variation over the past few centuries. We therefore conclude that the fact that LD wh-questions in contemporary corpora show such limited variation is due to a number of independent factors, and not a priori caused by the fact that LD wh-questions are formed by analogy to a general template.

References

- Ambridge, Ben & Adele Goldberg 2008. "The island status of clausal complements: evidence in favor of an information structure explanation." *Cognitive Linguistics* 19.349-381.
- Andersson, Sven-Gunnar & Sigmund Kvam 1984. *Satzverschränkung im heutigen Deutsch*. Tübingen: Narr.
- Behaghel, Otto 1928. *Deutsche Syntax. Eine geschichtliche Darstellung. Vol III: Die Satzgebilde*. Heidelberg: Winter.
- Blatz, Friedrich 1896. *Neuhochdeutsche Grammatik*. Karlsruhe: Lang.
- Chomsky, Noam 1977. "On Wh-Movement". *Formal Syntax* ed. by Peter Culicover, Thomas Wasow & Adrian Akamajian. Academic Press: New York.
- Dąbrowska, Ewa 2004. *Language, Mind and Brain*. Georgetown: Georgetown University Press.
- Dąbrowska, Ewa 2008. "Questions with long-distance dependencies: A usage-based perspective". *Cognitive Linguistics* 19.391-425.
- Ebert, Robert P. 1973. "On the Notion 'Subordinate Clause' in Standard German". *You take the High Node and I'll take the Low Node: Papers from the Comparative Syntax Festival*, 164-177. Chicago: Chicago Linguistic Society.
- Hoeksema, Jack & Ankelien Schippers 2009. "Diachronic changes in long-distance dependencies: the case of Dutch." *Submitted*.
- Howe, Stephen 1996. *The personal pronouns in the Germanic languages*. Berlin: De Gruyter.
- Kvam, Sigmund 1983. *Linksverschachtelung im Deutschen und Norwegischen. Eine Kontrastive Untersuchung zur Satzverschränkung und Infinitivverschränkung in der deutschen und norwegischen Gegenwartssprache*. Tübingen, Max Niemeyer Verlag.
- Lühr, Rosemarie 1988. "Zur Satzverschränkung im heutigen Deutsch." *GAGL* 28.74-87.
- Paul, Hermann 1920. *Deutsche Grammatik IV*. Halle: Max Niemeyer Verlag.
- Salzmann, Martin 2006. Resumptive Prolepsis. A Study in Indirect A' dependencies. PhD diss., University of Utrecht.
- Schippers, Ankelien 2009. "On the (un)availability of long-distance movement." To appear in: *Movement and Clitics* ed. by Vincent Torrens, Linda Escobar, Anna Gavarró & Junkal Gutierrez-Mangado. Newcastle: Cambridge Scholars Publishing.
- Verhagen, Arie 2005. *Constructions of Intersubjectivity*. Oxford: Oxford University Press.
- Verhagen, Arie 2006. "On subjectivity and 'long distance Wh-movement'." *Subjectification: Various Paths to Subjectivity* ed. by Angeliki Athanasiadou, Costas Canakis & Bert Cornillie, 323 - 346. Berlin/New York: Mouton de Gruyter.