Relative Clause Extraposition in Low Saxon

Jan Strunk

June 24, 2004

1 Introduction

This paper is a small corpus study of Low Saxon relative clause extraposition. It tests the predictions concerning this phenomenon made in Hawkins (2004).

Hawkins (2004) proposes three different principles to account for a large variety of language internal and typological variation. The principle that plays the most important role in word order variation is Minimize Domains (MiD).

> The human processor prefers to minimize the connected sequence of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization difference in each domain. (Hawkins, 2004, p. 205)

The most interesting cases in which this principle applies are those in which there is a competition between the minimization preferences of different phrasal combination domains. One such case that Hawkins discusses in his book is the phenomenon of relative clause extraposition (Hawkins, 2004, chapter 5.5).

Relative clauses are usually quite long and therefore lead to a longer phrasal combination domain between the head noun of a verb's NP argument and the verb itself if the relative clause intervenes between the nominal head and the verb. In such languages as German, Dutch, and the language that I will deal with here Low Saxon, the relative clause follows the head noun and can intervene between the nominal head and the verb in case the verb appears clause finally; cf. example (1).

(1) *Noch nienich heff ik* ***enen Minschen****, [de so innig*
yet never have I a human being who so intimately
mit een Bloom tosamenwuß], seihn.
with a flower grew together seen
"I have never seen a human being who grew together with a flower so intimately."

The verb is *seihn* (seen). I have marked the relative clause with square brackets and the head noun with boldface. In example (1), a seven word relative clause is located between the clause-final verb and the head noun of its object. Example (1) is actually a modified version of the authentic example (2) in which the relative clause has been extraposed after the verb, yielding a much smaller domain between the head noun of the object of the verb and only a slightly longer domain between the head noun and its relative clause in comparison to example (1).

(2) *Noch nienich heff ik* ***enen Minschen*** *seihn, [de so innig mit een Bloom tosamenwuß].*
yet never have I a human being seen who so intimately with a flower grew together
"I have never seen a human being who grew together with a flower so intimately."

The following examples give the general structure of the domains that I will be talking about in this paper.

1. Non-Extraposed: $[_{NP}$ enen Minschen$]$ $[_{RC}$ de so innig mit een Bloom tosamenwuß$]$ $[_{V}$ seihn$]$
2. Extraposed: $[_{NP}$ enen Minschen$]$ $[_{XP}$ $[_{V}$ seihn$]]$ $[_{RC}$ de so innig mit een Bloom tosamenwuß$]$

I will abbreviate the relative clause with RC and all the material that possibly intervenes between the NP and its relative clause as XP (this also includes a V if it comes between the NP and its relative clause).

The following are my interpretation of the detailed predictions that Hawkins (2004, pp. 274–275) makes for relative clause extraposition in German:

1. If the length of the XP between an NP and its relative clause is only one word long, extraposition will always be preferred.
2. Extraposition will get more and more dispreferred as the length of the XP increases.
3. The longer the relative clause, the more preferred will the extraposed variant be.

Hawkins cites a study by Uszkoreit et al. (1998) which supports his theory of relative clause extraposition with German data. In the following sections, I will first introduce my corpus and method of data annotation and then report my results and compare them to the results on German by Uszkoreit et al. (1998).

2 The Corpus

I have used a smaller section of a Low Saxon corpus that I have collected from the Internet for a computational linguistic paper (Strunk, 2003). The corpus

that I have used in the present corpus study comprises 396 smaller online texts in the dialects of north western Germany with a total number of approx. 271,117 words. I searched for all relative clauses marked by one of the following relative pronouns: *dat*, *wat*, *de*, and *den*. This resulted in 1285 relative clauses.

However, one can only decide whether a relative clause is extraposed or not if there is a so-called *rechte Satzklammer* (right sentence bracket), i.e. if there is a clause final verb or verbal particle which marks the boundary of the clause, i.e. the non-extraposed material (cf. also Uszkoreit et al., 1998). Therefore I excluded all relative clauses like the one in example (3) for which it is not clear whether it has been extraposed or not from further consideration.

(3) *In't* *Seniorenheim* *wöörn* ***noog*** ***Lüüd*** *[de* *mit*
in=the home for elderly people were enough people who with
ehr *snacken* *wulln]*
her talk want
"In the home for elderly people, there were enough people who wanted to talk to her."

Excluding such examples yielded a final sample of 630 relative clauses. For these remaining examples, I automatically counted the length of the relative clause and the length of the XP material that possibly intervened between the NP and its relative clause after having bracketed them manually

3 Results

The results that I obtained bear out the predictions made by Hawkins (2004) and do not differ much from those reported for German in Uszkoreit et al. (1998).

335 (53 %) of all relative clauses are extraposed. 295 (47 %) are not extraposed but situated clause-internally.

Table (4) gives the average lengths of the relative clauses for the two cases: EXTA(posed) relative clause and NONEXTRA(posed) relative clause and the corresponding standard deviations. Furthermore, it also reports the average length of the XP for the EXTRA case and the average length of the XP that would have resulted if the relative clause had been extraposed for the NONEXTRA case.

(4) Length of relative clause (RC) and intervening material (XP)

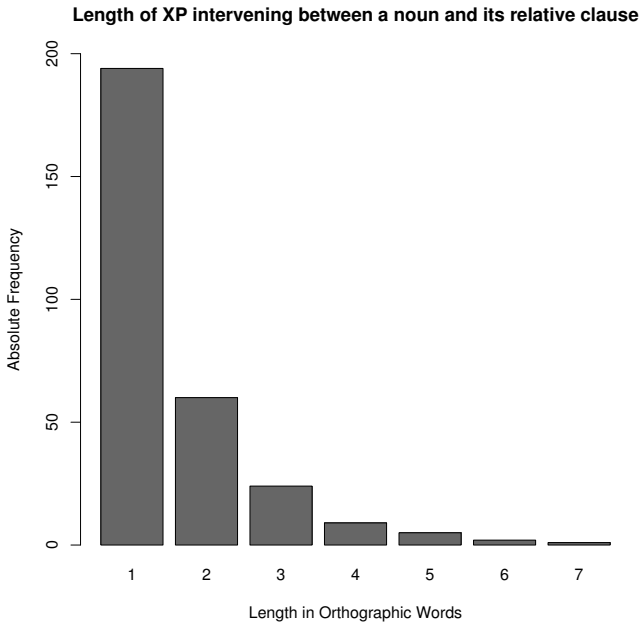
Type	**Mean RC Length**	**SD**	**Mean XP Length**	**SD**
EXTRA	7.49	4.07	1.58	1.02
NONEXTRA	6.26	2.93	(5.24)	(3.58)

The average length of the relative clause in the extraposed case is 1.23 words longer than that in the non-extraposed case. This difference is statistically

significant (t = 4.3103, df = 526.639, p = 1.946e-05). A similar difference of approx. 1 word is also reported for German in Uszkoreit et al. (1998). If we compare the average length of the actual XP in the EXTRA cases with that of the could-have-been XPs in the NONEXTRA cases, we see that the could-have-been XPs are significantly longer on average (t = -17.9363, df = 394.232, p =< 2.2e-16).

Just as Hawkins (2004) predicts and as Uszkoreit et al. (1998) have found, the length of the intervening XP is usually very small in the EXTRA cases. This tendency can also be seen in the number of extraposed relative clauses depending on the length of XP plotted in figure (5).

(5) Number of extraposed relative clauses

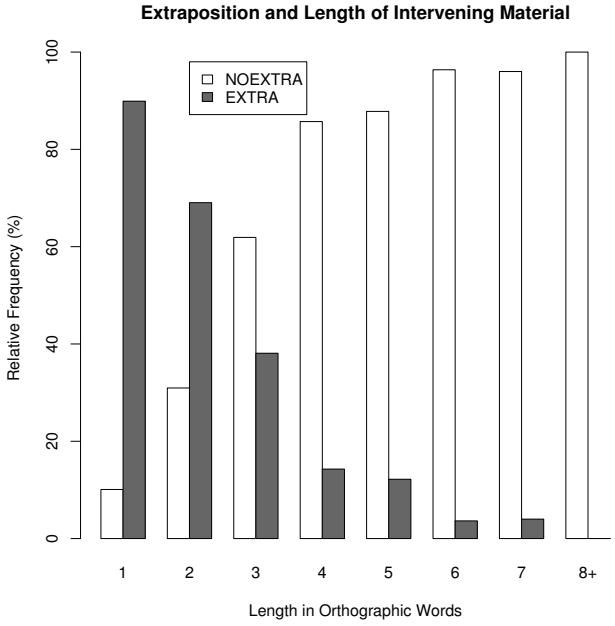


Most extraposed relative clauses are between one and three words long. The influence of the length of the intervening XP on the choice between an extraposed relative clause and a non-extraposed relative clause can be seen in figure (7). It shows that the longer the intervening material is that separates a relative clause from the NP it modifies the less likely it is that the relative clause will be extraposed. Table (6) gives the number of extraposed and non-extraposed relative clauses for all length of the intervening XP. Again the percentages closely resemble those obtained in a corpus study on written German by Uszkoreit et al. (1998).

(6) Relative clause extraposition depending on the length of the intervening XP

Type	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8+**
EXTRA	196	58	24	9	5	2	1	0
EXTRA (%)	89.91	69.05	38.10	14.29	12.20	3.64	4.00	0.00
NONEXTRA	22	26	39	54	36	53	24	63
NONEXTRA (%)	10.09	30.95	61.90	85.71	87.80	96.36	96.00	100.00

(7) Percentage of extraposed vs. non-extraposed relative clauses depending on the length of the XP



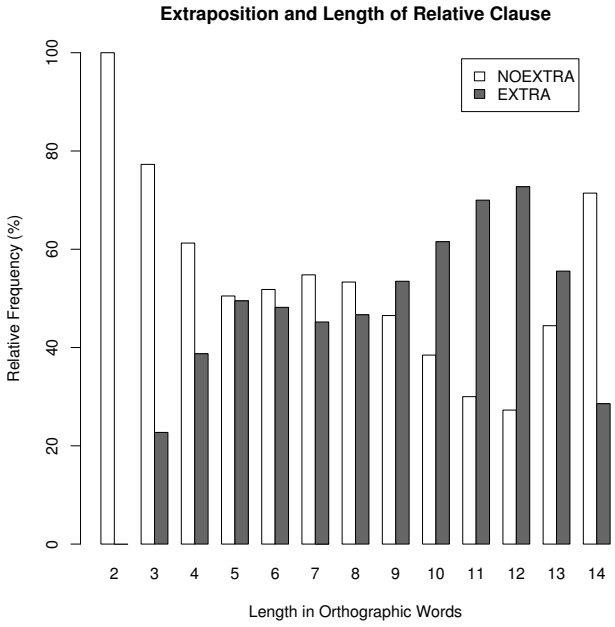
We thus see a very strong effect of the length of the material (possibly) intervening between the NP and the relative clause on the choice of whether to extrapose a relative clause or not. The effect of the length of the relative clause itself can also be seen in figure (9) but does not seem to be as important as the length of the intervening XP.[1]

[1] Figure (9) does not show the distribution for longer relative clauses which occur only very infrequently.

(8) Relative clause extraposition depending on the length of the relative clause

Type	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
EXTRA	0	10	43	50	49.50	53	33	21	23
EXTRA (%)	0.00	22.73	38.74	48.18	45.21	46.67	53.49		
NONEXTRA	5	34	68	51	57	40	24	20	
NONEXTRA (%)	100.00	77.27	61.26	50.5	51.82	54.79	53.33	46.51	
Type	**10**	**11**	**12**	**13**	**14**	**15+**			
EXTRA	16	14	8	5	2	16			
EXTRA (%)	61.54	70.00	72.73	55.56	28.57	66.67			
NONEXTRA	10	6	3	4	5	8			
NONEXTRA (%)	38.46	30.00	27.27	44.44	71.43	33.33			

(9) Percentage of extraposed vs. non-extraposed relative clauses depending on the length of the relative clause



I also conducted logistic regression with the variable *EXTRA/NONEXTRA* as dependent variable and the variables *length of the relative clause* and *length of XP* as predictor variables. I have also added the variable *length of the NP without the relative clause* to see whether it has an effect.

The output of this analysis in figure (10) shows that the length of the (possibly) intervening XP is clearly the most important factor in the decision whether to extrapose the relative clause or not. However, the length of relative clause itself still has a significant influence on this decision.

(10)
```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9604  -0.3784   0.1053   0.3502   1.1504

Coefficients:  Estimate    Std. Error  t value    Pr(>|t|)
(Intercept)     0.682192   0.046756     14.590     < 2e-16  ***
Length of RC    0.017531   0.004615      3.798     0.000160 ***
Length of XP   -0.083832   0.005029    -16.670     < 2e-16  ***
Length of NP   -0.018692   0.013289     -1.407     0.160048
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Null deviance: 156.87  on 629  degrees of freedom.

Residual deviance: 104.74  on 626  degrees of freedom

AIC: 667.5
```

The length of the NP (without the relative clause) does not have an effect which is also in line with Hawkins (2004) who did not predict any effect for this variable.

My results thus agree with both Hawkins' predictions and the results on written German by Uszkoreit et al. (1998).

4 Conclusion

I was able to show that the same effects operate in Low Saxon relative clause extraposition that are predicted by the *Minimize Domains* principle in Hawkins (2004) and have been demonstrated for German by Uszkoreit et al. (1998). These results are not too surprising given that German and Low Saxon have almost identical word order possibilities and are also otherwise closely related but still provide further evidence for Hawkins' theory.

During the annotation of the data, I found two interesting phenomena that could be relevant for the discussion in Hawkins (2004). First, it seems that Low Saxon (and German) do tolerate quite long NPs with relative clauses in the preverbal position in verb-second clauses; cf. example (11) (I have marked the verbs with angle brackets.).

(11) *Vun* ***de*** ***Lüüd*** *[de bi den Anslag vun den 11. September*
from the people who by the attack of the 11th September
ehr Neegsten verloren hebbt] <weer> to <hören>, dat se
their relatives lost have, was to hear that they
dat Ordeel meist nich glöven kunnen.
the verdict almost not believe could
"The people who have lost relatives in the attacks of September 11th stated that they almost could not believe the verdict."

The relative clause from such long NPs would usually be extraposed if the NP were situated in the Mittelfeld (middle field). There could be multiple reasons for allowing such long NPs in preverbal position. The length of the intervening XP could simply be too long if one would extrapose the relative clause. There could be other factors such as the presence of a long complement after the verb as in example (11) or simply information structure which overrides length differences. It might also be the case that the reason why long NPs in the first position in the clause are not so bad is that one has not encountered the constructing category of the VP yet and simply does not take the long NP into account in the calculations for domain minimization (Hawkins, 2004, chapter 5).

Second, as there is no standard written variety of Low Saxon and writing Low Saxon is not taught in school, Low Saxon authors often write in a style that is closer to the spoken language and therefore also use sentences that would probably not be considered grammatical according to the grammar rules taught in school.

In a lot of sentences like the one given in example (12) which contain a sentence initial NP with a long relative clause, a different strategy from relative clause extraposition is chosen. The initial NP is resumed by a demonstrative pronoun. I have found 58 such examples.

(12) ***Vele*** ***Straten*** *[de den Barg dal loopt na't Water*
many streets that the hill down run to=the water
hen] ***dat*** *sünd "Gruben".*
DEICT that are pits
"Many streets that come down the hill they are called 'Gruben' (pits)."

I would venture to say that the "left-dislocation" in these examples is not parallel to ordinary left-dislocation for discourse functional reasons but probably simply due to the length of the initial NP. Such examples could therefore be interesting as an alternative strategy to relative clause extraposition.

References

Hawkins, J. A. (2004). *Efficiency and complexity in grammars.* Oxford University Press, Oxford.

Strunk, J. (2003). Information retrieval for languages that lack a fixed orthography. Term paper, Stanford University, http://www.stanford.edu/∼jstrunk/.

Uszkoreit, H., T. Brants, D. Duchier, B. Krenn, L. Konieczny, S. Oepen, and W. Skut (1998). Studien zur performanzorientierten Linguistik: Aspekte der Relativsatzextraposition im Deutschen. *Kognitionswissenschaft 7*, 129–133.