The Information Structure of Subject Extraposition in Early New High German

Caitlin Light*

1 Introduction

This paper investigates the information-structural characteristics of extraposed subjects in Early New High German (ENHG). Based on new quantitative data from a parsed corpus of ENHG, I will argue that unlike objects, subjects in ENHG have two motivations for extraposing. First, subjects may extrapose in order to receive narrow focus, which is the pattern Bies (1996) has shown for object extraposition in ENHG. Secondly, however, subjects may extrapose in order to receive a default sentence accent, which is most visible in the case of presentational constructions. This motivation does not affect objects, which may achieve the same prosodic goal without having to extrapose.

I will begin from the following information-structural assumptions. This paper is roughly based on a theory of information structure as in Vallduví (1992), in which every sentence is divided into a Focus-Ground partition, and every sentence has only one information-structural focus. I assume that any constituent (and possibly some non-constituents) may be the focus of a sentence. I use the term *narrow focus* to describe a DP which is, in itself, the sole focus of a clause (that is, neither part of the Ground, nor part of a larger focused constituent). The term *narrow focus* in this sense covers a number of more specific focus types.

I will also use the term *presentational focus*, which describes the type of focus which introduces a new entity into the discourse (it has in other works been described by the term *new information focus*). As I will discuss later, a newly introduced DP in a presentational construction may be narrowly focused, or it may not. In other contexts, narrowly focused DPs may also be interpreted as other types of focus, for example contrastive focus, which will not be discussed in detail in the current paper.

The remainder of this paper will be organized as follows. In the next section, Section 2, I will discuss some previous studies on related phenomena, which will help to structure the current investigation. Then, in Section 3, I will outline the methodology and quantitative results of the current study. I propose an analysis of these results in Section 4; finally, in Section 5, I offer some concluding remarks.

2 Related Studies

Although I know of no previous studies of subject extraposition in ENHG (particularly from a quantitative perspective), some related phenomena in Germanic have been explored by previous works. For my purposes, the most notable are two studies based on quantitative data: Bies (1996) on object extraposition in ENHG, and Prince (1989) on subject extraposition in Yiddish. A brief review of these two works is necessary before proceeding to the current study.

2.1 Object Extraposition in ENHG

Bies (1996) provides a detailed analysis of the information structure of DP extraposition, based on a corpus of examples collected from various ENHG texts. She considers two possible motivations for extraposition, discourse newness and narrow focus on the DP. Before considering discourse factors, however, Bies identifies external influences on DP extraposition; first, quantified and indefinite

^{*}This paper will also be appearing in the proceedings of the HPSG 2011 Conference. I am very grateful to Anthony Kroch, Joel Wallenberg, Ivona Kucerova and Aaron Ecay for their help and advice on various stages of the development of this paper. Thanks also to the attenders of the 35th Annual Penn Lingustics Colloquium (PLC) and the Information Structure in Formal Grammar (IFG) workshop for their comments and feedback. All errors are, naturally, my own.

	Postposed	Non-post.	Rate of post.
Negation	0	86	0%
Indefinite	8	301	2.3%
Other QP	2	83	2.4%
Non-quant.	64	408	13.6%

objects are much less likely to extrapose (Table 1). Second, DP length (or 'weight') also strongly influences extraposition (Table 2).

Table 1: DP type of extraposed objects in ENHG (Bies 1996).

	Postposed	Non-post.	Rate of post.
Conjoined	13	17	43.3%
Relative clause	5	2	71.1%
PP postmodifier	25	34	42.4%
Regular length	82	722	10.2%
Total	125	775	13.9%

Table 2: Modifiers as a measure of DP weight for extraposed objects in ENHG (Bies 1996).

Based on these observations, Bies restricts her data set to non-quantified DPs of 'regular length' (that is, without PP modifiers, relative clauses or conjunction). She also excludes topicalized and scrambled objects from the set of non-extraposed DPs, assuming that they represent unrelated information structural phenomena. Bies then separates her (restricted) data set into three informational categories: discourse-new, evoked/inferred, and given information, adapted from a broader hierarchy of information types in Prince (1981). She observes a gradient relationship between newness and extraposition (Table 3).

	Postposed	Non-postposed	Rate of postposing
Given	11	100	10%
Evoked/Inf.	37	81	31.4%
Discnew	16	21	43.2%
Total	64	202	24.1%

Table 3: Discourse status of extraposed objects in ENHG (Bies 1996).

Bies then asks: is this the main discourse motivation for DP extraposition, or a symptom of it? She suggests that discourse-newness of a DP may contribute to its likelihood of being narrowly focused: discourse-new elements often are the focus of a sentence. Perhaps the relationship between discourse-newness and extraposition is simply a consequence of the fact that these elements are more likely to be narrowly focused. To explore this alternate hypothesis, she further classifies her sentences into (narrow) DP focus and (wide) VP focus, wherever context allows an unambiguous classification (Table 4). The effect of narrow focus, as the data show, is stronger than that of information status. Bies therefore concludes that narrow focus alone motivates object extraposition in ENHG.

2.2 Subject Extraposition in Yiddish

One could simply assume that Bies's conclusions may naturally extend as the general pattern for DP extraposition in ENHG, covering both subjects and objects. However, a second study suggests that another analysis may be possible. There is some reason to believe that subject extraposition may

	DP focus	VP focus	Percent DP focus
Non-postposed DP	19	123	13.4%
Postposed DP	46	4	92%

Table 4: Focus structure of clauses with an extraposed object in ENHG (Bies 1996).

occur for independent purposes within a related Germanic language. Prince (1989) considers the pragmatic properties of subject extraposition in Yiddish, a language closely related to ENHG, and argues that in this case, the extraposition is motivated by the discourse-new status of the DP.

Prince begins by considering a set of examples collected from a Yiddish text. She finds that in certain subordinate clause types, brand-new subjects are highly motivated to postpose (Table 5). This leads her to suggest that discourse status is strongly related to the motivation for subject extraposition in Yiddish.

	Non-postposed	Postposed	Total
Adverbial	16 (57%)	12 (43%)	28
Complement	14 (30%)	32 (70%)	46
Total A/C	30 (41%)	44 (59%)	74
Free Relative	0 (0%)	7 (100%)	7
Indirect Q.	0 (0%)	3 (100%)	3
Relative	0 (0%)	2 (100%)	2
Total WH	0 (0%)	12 (100%)	12

Table 5: Extraposition of brand-new subjects in Yiddish by clause type (Prince 1989).

Prince also proposes a syntactic motivation for her analysis. Although the data are small, she notes that extraposition of discourse-new subjects appears to be categorical in relative clauses, free relatives, and indirect questions. She links this to another phenomenon seen in Yiddish: the expletive *es* ('it') is licensed to fill Spec,TP when the subject is extracted from a free relative or indirect question (Prince 1989, Diesing 1990). Crucially, this does not occur in subject relative clauses.

- (1) a. *Ikh veys nit ver es iz gekumen* I know not who ES is come
 - 'I don't know who came.'
 - b. *Ikh veys nit ver iz gekumen
- (2) a. *Der melamed vos iz besser far ir iz beser far mir.* the teacher that is better for her is better for me 'The teacher that is better for her, is better for me.'
 - b. *Der melamed vos es iz beser far ir iz beser far mir.

Prince hypothesizes that this expletive appears when the subject is extracted from a postposed position, leaving Spec,TP empty. This would provide a pragmatic explanation for the fact that the expletive appears only in indirect questions and free relatives, which (unlike relative clauses) typically have an extracted element that is discourse-new.

Prince therefore provides both quantitative and syntactic arguments for the claim that subject extraposition in Yiddish is motivated by the discourse status of the subject. This raises a question: should ENHG have a unified analysis for DP extraposition as a whole, or do subject and object extraposition behave differently? Phrased differently, do subjects in ENHG pattern like subjects in a related language, or like non-subject DPs in ENHG? My goal in this study is to propose an answer to this question, based on new data on subject extraposition in ENHG.

3 The Current Study

In this section, I will outline the methodology for the current study, and present the quantitative results. This will set the stage for Section 4, in which I present an analysis of subject extraposition in ENHG. I will argue that, although narrow focus proves to be a factor in both subject and object extraposition in ENHG, there is an additional motivation for subject extrapositio. This is in fact related to the discourse status of the subject, as Prince (1989) argued for Yiddish.

3.1 Methodology

The data for my study were drawn from a parsed corpus of Martin Luther's first New Testament translation, the *Septembertestament*, published 1522. The *Septembertestament* corpus consists of rougly 102,000 words, including the full text of Matthew, Mark, John and the Acts of the Apostles. It was initially parsed by automatic methods (including Bikel 2004), but ultimately I hand-corrected the full text. The parsing format is modeled on the guidelines for the Penn Historical Corpora of English and the York-Toronto-Helsinki Parsed Corpus of Old English Prose (cf. Kroch and Taylor 2000, Kroch et al. 2004, Taylor et al. 2003), adapted for use for a German corpus.

From the Septembertestament corpus, I extracted all unambiguous examples of extraposed subjects, using clause-final verbs and verbal particles as diagnostics. I found 115 examples which fit these requirements. All were hand-coded for definiteness, syllable length, and discourse status of the subject (based on Bies's classifications). I also coded the examples for focus structure using a binary measure: either (1) they had narrow focus on the extraposed subject or (2) they had a focus structure of another type (I did not code in more detail in this case). These were compared to 1261 examples of subjects that could have been extraposed but were not. This excludes pronominal subjects (including impersonal *man*, 'one') and demonstrative determiners, which are too light to extrapose. Each of these tokens was coded for syllable length and the definiteness of the subject. Different subsamples of this set were isolated for the consideration of the pragmatic and information-structural characteristics of subject extraposition, which will be discussed shortly.

3.2 Quantitative Results

In the sample collected, subject extraposition occurs at an overall rate of 8.4%. This is lower than the rate of object extraposition described in Bies (1996), 13.2%.

As Bies showed for object extraposition, weight proved to be a strong influence on subject extraposition (I deviate from Bies in measuring DP weight by syllables, rather than by modifier presence and type). The minimum weight of an extraposed subject was 2 syllables, while the maximum was 64 (due to a sequence of embedded clauses within the DP). The average weight of extraposed subjects was 13.07 syllables. In comparison, the minimum weight of a non-extraposed subject was 1 syllable, and he maximum was 29. The average weight of non-extraposed subjects was only 3.29 syllables.

To minimize the effect of DP weight on the sample, I chose to limit my sample to subjects of 15 syllables or less. This ensures that the DPs are of a weight safely below the limit found on non-extraposed subjects, without too greatly restricting the data set. The remainder of the paper deals only with this subset of the data, unless otherwise noted. The adjusted sample includes 86 extraposed subjects and 1257 non-extraposed subjects, or extraposition at a rate of about 6.4%.

The consideration of definiteness exposes a striking difference between subject and object extraposition: quantified/numeric subjects are extraposed more frequently than definites. This is true of the entire sample regardless of weight; Table 6 shows the distribution of subject types for the full sample of clauses.

Note, however, that the sample contains no extraposed negated subjects. This is an interesting fact, but a detailed discussion is beyond the scope of the current paper. Furthermore, bare subjects extrapose more often than not. It is unclear how this relates to Bies's data on object extraposition, because her examples suggest that she may have included bare DPs in her 'non-quantified' category

	Non-extraposed	Extraposed	% Extraposed
Negative	47	0	0.0%
QP/Numeric	76	17	18.28%
Indefinite	98	9	8.41%
Bare	12	13	52.00%
Free rel.	13	28	68.29%
Definite	1015	48	4.52%
Total	1261	115	8.36%

Table 6: DP types of extraposed subjects in ENHG.

(but this fact is never explicitly noted). Regardless, the overall distribution of DP types demonstrates a contrast between subject and object extraposition in ENHG.

I then compared the discourse status of all extraposed subjects to all non-extraposed subjects in a subset of 443 matrix and 173 subordinate clauses (for purposes of examining a smaller subset of the data); clauses with subjects of more than 15 syllables are excluded. As Table 7 demonstrates, the majority of extraposed subjects are discourse-new, but the majority of non-extraposed subjects are given. This is as expected, based on Bies's results for object extraposition in ENHG.

	Discourse-new	Evoked/Inf.	Given
Extraposed	33 (38.82%)	26 (30.59%)	26 (30.59%)
Non-extraposed	75 (12.17%)	165 (26.79%)	376 (61.04%)

Table 7: Discourse status of extraposed subjects in ENHG.

The 86 clauses with extraposed subjects were then compared to a randomly selected sample of 60 clauses with non-extraposed subjects, for a detailed consideration of the focus structures of these groups. Contextually and structurally ambiguous examples were set aside. An example of an extraposed subject coded for narrow focus is given in (3). Note that the extraposed subject contains the focus particle *auch*, making the focus structure particularly clear.

(3) denn es werden falsche Christi, vnd falsche propheten auff stehen, vnd grosse tzeychen vnd for it will false Christs and false prophets up stand and great signs and wunder thun das verfuret werden, yhn denn yrthum wo es muglich were auch wonders do that misled will.be in the confusion where it possible would.be also die auserweleten.

the chosen

'For false Christs and false prophets will come forward and perform great signs and wonders, so that in the confusion, where possible, even the chosen will be misled.' (*Septembertestament*, Matthew 24:24)

As discussed above, this study creates a binary distinction between narrow focus on the extraposed subject and any other focus structure, all of which are included in the "Other foc." category. I find that extraposed subjects are narrowly focused more often than non-extraposed subjects (Table 8). However, whereas Bies found that 92% of extraposed objects were narrowly focused, only 62% of subjects in my data are narrowly focused. This leads me to reject the hypothesis that subject extraposition is driven solely by narrow focus as a broad category.

The data discussed in this section suggest that the relationship between subject and object extraposition is not simple. In many cases, there is a distinct similarity between the two: both show an effect of both discourse status and focus structure, although to different degrees. However, there are also some distinct differences, most notably in the DP types associated with each: while object extraposition occurs more frequently with definite DPs, subject extraposition shows the opposite tendency.

	Narrow S-foc.	Other foc.	% Narrow foc.
Extraposed	31	19	62.0%
Non-extraposed	4	41	8.9%
Total	35	60	36.8%

Table 8: Focus structure of clauses with extraposed subjects in ENHG.

In the following section, I will present a proposal that may capture both the similarities and differences outlined above. I will suggest that the general motivation of subject extraposition is prosodic; while narrow focus is certainly the central motivation for some cases of subject extraposition, I will suggest that a different, more specific information structural phenomenon underlies the cases that define the difference between subject and object behavior.

4 Analysis of Subject Extraposition

As the data in the preceding section suggest, there are some ways in which subject and object focus pattern similarly. For example, many examples can be found which involve narrow focus on an extraposed definite subject, as (4) demonstrates (note that two contrastive subjects have been extraposed in two separate clauses). This is exactly the sort of example we expect if subject and object extraposition share the same information structural properties in ENHG. I argue that in cases such as these, subject and object extraposition do in fact have the same motivation: narrow focus on the DP.

(4) vnnd eynem gab er funff centner, dem andern zween, dem dritten eyn, eynem ydern noch and one gave he five talents the other two the third one one each after sevnem vermugen vnd zoch hynweg ... vnd da tratt ertzu, der da funff centner his ability and went away ... and then tread forward who PART five talents empfangen hatte ... Do trat auch ertzu, der do zween centner empfangen hatte received had ... Then tread also forward who PART two talents received had . . .

'And he gave five talents to one, two to another, one to the third, each according to his ability, and went away ... and then the man who had received five talents came forward' (Septembertestament, Matthew 25:15–22)

However, the frequencies at which quantified subjects extrapose suggests a similarity between subject extraposition in ENHG and subject extraposition in Yiddish, where the discourse status of the subject plays a more important role. I assume that indefinite and quantified DPs are more likely to be discourse-new. The higher rate of extraposition of indefinite/quantified subjects suggests a stronger effect of discourse status on subject than on object extraposition in ENHG. Based on this fact, I will argue that subject extraposition in ENHG can also be motivated by a more specific type of focus.

In German, the sentence accent generally falls on the rightmost argument of the VP by default, even if followed by a clause-final non-finite verb or verbal particle (see Ladd 1996, Truckenbrodt 2007: for a summary of the literature on this). As a result, object DPs *in situ* are frequently in the appropriate position to receive default sentence accent. Scrambling of other elements can further help to situate a non-topicalized DP at the right edge of the 'middle field,' so that it may be in the rightmost position and receive the default accent when necessary. Because this is permitted, extraposition of object DPs may be expected to have more specific motivations than simply to obtain default accent. This seems to be compatible with Bies's analysis. However, more elaborate means are often required to maneuver the subject into the location of default accent. For example, expletive *es* may be inserted in topic position, while the subject appears in a low position.

I propose that subject extraposition may be used as a general means to obtain a default accent on a subject, without resulting in a contrastive interpretation. There is a specific clause type that may demonstrate this: clauses with presentational focus. In fact, a large subset of the extraposed subjects are presentational, and best translated into English with the use of existential *there*, as in (5).

(5) Aber die kinder des reychs werden außgestossen ynn die außersten finsternisß, da but the children of-the kingdom will-be cast-out in the outermost darkness there wirt seyn weynen vnd tzeen klappen.
will be weeping and teeth gnashing 'But the children of the kingdom will be cast out into the outermost darkness. In that place there will be wailing and gnashing of teeth.' (Septembertestament, Matthew 8:12)

Other clauses may be found which have the same effect of introducing a new entity into the discourse, but do not appear to involve narrow focus on the extraposed DP (but rather a broader focus structure). I present (6) as a clear example of this phenomenon; note that the modifiers on the subject are intended to introduce the entity it denotates, emphasizing the fact that the entity has not been previously discussed.¹ However, the focus structure of the clause is broad, as the remainder of the information (namely, the actions taken by the newly introduced Gamaliel) is also newly introduced and focal.

(6) Da stund aber auffym radt eyn Phariseer mit namen Gamaliel, eyn schrifftgelerter, then stood however up in-the council a Pharisee with name Gamaliel a scholar wolgehallten fur allem volck well-held before all people
'But then a Parisee named Gamaliel stood up in the council, a scholar, well regarded by all the people.'
(Septembertestament, Acts 5:35)

The nature of these subjects implies that they will generally be quantified, indefinite or bare DPs. Additionally, they will generally be new entities in the discourse. This embodies the difference between subject and object extraposition in ENHG. We can also observe this effect by considering the occurrence of copular clauses in each data set: while 16 (18.6%) of extraposed subjects occur in copular clauses, only 92 (7.3%) of non-extraposed subjects do. This means that copular clauses extrapose at a rate of 14.8%, while non-copular clauses extrapose at a rate of 5.67%. My argument is that the link between subject extraposition is due to the fact that many copular clauses are presentational, and thus favor extraposition of the subject above other clause types.

4.1 Extraposition and Sentence Accent

Before concluding, I will offer some brief remarks on the relation between extraposition and sentence accent. Ladd (1996) presents a metrical account of sentence accent. He observes that the accent patterns of the two sentences in (7) may differ, even when both are interpreted with broad (sentential) focus. He proposes that this can be explained by the fact that the shorter utterance may consist of only one intermediate intonational phrase, and within this intermediate phrase the primary accent falls on the subject. However, once the utterance is as long as it is in (7b), the subject and predicate may not form a single intonational phrase, and must be split into two intermediate phrases. These two phrases have a weak-strong relation, so that the primary sentence accent falls on the strongest accent in the second intonational phrase. Ladd notes, "The heavier a constituent is, the more likely it is to constitute its own intermediate phrase."

(7) a. JOHNSON died.

¹However, the length of this subject would make it a non-ideal example otherwise; I will simply mention that, while this is a particularly useful example for demonstrating the phenomenon under consideration, other examples exist which do not have its shortcomings.

b. Former president Johnson unexpectedly DIED today.

Wallenberg (p.c.) proposes that in English Heavy NP-Shift (HNPS), a "Heavy NP" moves rightward past any material on its right in order to consitute its own (rightmost) intermediate phrase, and thus bear the primary sentence accent. I tentatively propose the same analysis for subject extraposition in ENHG: in order to receive the primary sentence accent while allowing a presentational (and non-contrastive) focus interpretation, the subject may move to the right edge. By extraposing, the subject forms its own intermediate intonational phrase, which enters a weak-strong metrical relation with other intermediate phrases in the sentence, and ultimately receives the primary sentence accent.

This proposal requires further testing, but it may help explain why subjects might extrapose in presentational contexts. Under this analysis, a central motivation of subject extraposition is prosodic. It may also help to explain why extraposition targets particularly heavy DPs: these are the DPs, in Ladd's own observation, which are most likely to constitute their own intermediate phrase. Extraposition may be a way to syntactically facilitate this.

5 Conclusion

I have argued, based on quantitative data from a parsed corpus of ENHG, that there are both similarities and differences between subject and object extraposition in ENHG. Both subjects and objects may be extraposed to express narrow focus on the extraposed DP. However, subjects may also be extraposed for a more specific motivation: as a means to achieve default accent on the subject, particularly in presentational contexts. As a result, subject extraposition occurs more frequently with quantified subjects, as well as with entities are new to the discourse.

The result of this conclusion is twofold. First, I have argued that subjects may have multiple motivations to extrapose: either to express narrow focus or to obtain a default sentence accent in a non-contrastive context. Second, I have shown that the phenomenon of DP extraposition in ENHG demonstrates a subject-object asymmetry.

This has certain consequences for the broader study of information structure. On one hand, subject extraposition in ENHG provides new evidence that there is not necessarily a one-to-one correspondence between syntactic construction and information structural interpretation; rather, in this case, the syntax may be manipulated to accomplish multiple information structural and prosodic goals (cf. Féry 2007). On the other hand, DP extraposition in ENHG fits into a broader set of crosslinguistic focus phenomena which demonstrate a subject-object asymmetry (cf. Hartmann and Zimmermann 2007, Skopeteas and Fanselow 2010), raising important questions about the relationship between argument structure and information structural notions.

On an unrelated note, I offer this study as a demonstration that parsed corpora may be used as resources in information structural research, and of the importance of quantitative data when exploring such subtle and complex issues. As the study of information structure progresses, and judgments become more elaborate and less reliable, large corpora of attested examples embedded within concrete contexts may become an important and valuable resource, offering a type of data that judgments alone cannot achieve.

References

- Bies, Ann. 1996. Syntax and discourse factors in Early New High German: Evidence for verb-final word order. Master's thesis, University of Pennsylvania.
- Bikel, Daniel M. 2004. On the parameter space of lexicalized statistical parsing models. Doctoral Dissertation, Department of Computer and Information Sciences, University of Pennsylvania.
- Diesing, Molly. 1990. Verb movement and the subject position in Yiddish. *Natural Language & Linguistic Theory* 8:41–79.
- Féry, Caroline. 2007. Information structural notions and the fallacy of invariant correlates. *Interdisciplinary studies on information structure* 6.
- Hartmann, K., and M. Zimmermann. 2007. In place out of place? Focus in Hausa. In On Information Structure: Meaning and Form, ed. K. Schwabe and S. Winkler, 365–403. Amsterdam: John Benjamins.

- Kroch, Anthony, Beatrice Santorini, and Ariel Diertani. 2004. Penn-Helsinki Parsed Corpus of Early Modern English. http://www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-2/index.html.
- Kroch, Anthony, and Ann Taylor. 2000. Penn-Helsinki Parsed Corpus of Middle English, second edition. http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-3/index.html.
- Ladd, D. Robert. 1996. *Intonational phonology*, volume 79 of *Cambridge studies in linguistics*. United Kingdom: Cambridge University Press.
- Prince, Ellen. 1981. Toward a taxonomy of given-new information. In *Radical pragmatics*, ed. P. Cole, 223–55. New York: Academic Press.
- Prince, Ellen. 1989. Yiddish Wh-clauses, subject-postposing, and topicalization. In *ESCOL* '88, ed. J. Powers and K. de Jong, 403–15. Columbus, OH: Ohio State University.
- Skopeteas, S., and G. Fanselow. 2010. Focus types and argument asymmetries: a cross-linguistic study in language production. In *Contrastive information structure*, ed. C. Breul and E. Göbbel, 169–197. Benjamins.
- Taylor, Ann, Anthony Warner, Susan Pintzuk, and Frank Beths. 2003. The York-Toronto-Helsinki Parsed Corpus of Old English Prose. University of York.
- Truckenbrodt, Hubert. 2007. The syntax-phonology interface. *The Cambridge handbook of phonology* 435–456.

Vallduví, Enric. 1992. The informational component. New York: Garland.

Department of Linguistics University of Pennsylvania Philadelphia, PA 19104–6305 clight@babel.ling.upenn.edu