

Swantje Tönnis, Lea M. Fricke, Alexander Schreiber

Methodological Considerations on Testing Argument Asymmetry in German Cleft Sentences

Abstract We present a corpus study on German *es*-clefts that tests whether subject clefts are more frequent than object clefts. This observation has been made for several other languages. However, we use a more complex method than earlier studies by not only providing the frequencies of subject/object clefts but by additionally comparing those frequencies to the general frequency of subjects/objects. Our results support the claim that subject clefts are more frequent in German. We argue that a cleft construction in its function to mark focus appears more often with subjects since there are additional options to mark focus on objects. Other features such as exhaustivity and contrast do not play a role in our cleft sample. From these results, we conclude that subjecthood is the main factor that facilitates the use of a cleft, possibly as a result of the author's intention to disambiguate focus.

Keywords German *es*-cleft, prosodic prominence, focus marking, argument asymmetry

1 Introduction

This paper presents a corpus study with the aim of contributing to a better understanding of the factors that facilitate the use of *es*-clefts in German. We analyzed crucial properties of clefts and their contexts. In this paper, we focus mainly on one aspect, namely the grammatical role of the pivot. Depending on the grammatical role of the pivot in the relative clause, we distinguish between subject clefts as in (1), and object clefts as in (2).

- (1) Es war Alt-Bundespräsident Roman Herzog, der zum
 It was former president Roman Herzog, who_{NOM.SG} on the
 50-jährigen Jubiläum eine internationale Neuorientierung
 50th anniversary a international re-orientation
 der Stiftung anregte.
 of the foundation suggested.
 'It was the former president Roman Herzog who suggested an interna-
 tional re-orientation of the foundation on the 50th anniversary.'
 (Z07/JUL.00590 Die Zeit [Online-Ausgabe], 19.07.2007; Noble Töne, enttäuschter
 Nachwuchs)

- (2) Es ist der Aufsteiger, den Balzac mit immer neuen
 It is the climber, who_{ACC.SG} Balzac with constantly new
 charakterlichen Merkmalen porträtiert, [...].
 character features portrays, [...].
 'It is the (social) climber who Balzac portrays with constantly new charac-
 ter features.'
 (R99/MAI.38158 Frankfurter Rundschau, 15.05.1999, S. 3, Ressort: ZEIT UND BILD;
 Zum 200. Geburtstag von Honoré de Balzac)

For several languages, it has been claimed that subject clefts are more frequent than object clefts (Carter-Thomas 2009, Roland et al. 2007, and Skopeteas & Fanselow 2010). We tested this claim for German clefts given that to our knowledge this has not been explicitly tested. Additionally, we use a more fine-grained method than earlier studies on other languages. We do not only provide the frequencies of subject and object clefts but also compare those frequencies to the general frequency of subjects and objects. It is important to take this additional step since it could be possible that subjects are just clefted more often because they are generally more frequent.

2 Background

The observation that subject clefts are more frequent than object clefts is closely related to focus marking. The cleft construction is one option for a language to realize focus, in addition to prosodic prominence, movement, and morphology. In some languages, not all of these options are equally available for all grammatical functions (see Lambrecht 2001 for French, or Hartmann & Zimmermann 2007 for West Chadic Languages). In French, for example, focus on objects can be realized via prosodic prominence, while this is not an option for subjects. According to Féry (2001), prosodic prominence is obligatorily realized at the right edge of the phonological phrase in French. Objects occur in this position and receive prosodic

prominence. Subjects, in contrast, cannot appear there. In the pivot of a cleft, however, subjects are located at the edge of a phonological phrase and receive default high prominence (see also Reinhart 1995: 62). The default intonation of a focus-background cleft¹ in French is exemplified in (3), taken from Destruel (2012).

- (3) C'est BATMAN qui a pour mission d'attraper les cambrioleurs.
 it-is BATMAN who has for mission to-catch the thieves.
 'It is Batman who has the mission of catching thieves.'

Accordingly, Szendrői (1999: 553) proposes to analyze clefts as focus-driven movement. Similarly, DeVeaugh-Geiss et al. (2015: 386) call clefts a structural device to mark focus unambiguously. Focus on an object NP can also be realized by a cleft construction. However, there are other options for focus-marking on objects (that are inapplicable to subjects), such as default intonation and scrambling. Hence, object NPs are predicted to be clefted less often than subjects.

The aim of our study is to analyze German data with respect to the frequency of subject and object clefts and thereby gain a deeper understanding of the function of a cleft sentence. More precisely, we discuss whether the primary function of a cleft is to mark focus. German, just as French, assigns the default accent at the edge of a phonological phrase. However, it allows for more variation when it comes to intonation (see Section 4 for a detailed discussion).

3 Corpus Study

3.1 Method

We drew a random sample of 300 clefts from a sub-corpus of the DeReKo corpus² of written German. In our annotation, we focused on well-defined properties like the grammatical function of the cleft relative pronoun,³ and the thematic role and animacy of the pivot NP. In order to account for the general frequency of

- 1 We will ignore topic-comment clefts in this paper, given that we found much more focus-background clefts in our corpus search.
- 2 Das Deutsche Referenzkorpus DeReKo (<http://www.ids-mannheim.de/kl/projekte/korpora/>), Institut für Deutsche Sprache, Mannheim.
 Since the annotation of some of the properties required a lot of context before and after the cleft sentence, we excluded texts that were not fully accessible. Moreover, we excluded *Wikipedia* articles because text coherence cannot be guaranteed due to possibly different authors for adjacent paragraphs of a text.
- 3 We only considered subject and object clefts. We did find some adjunct clefts where the relative pronoun was preceded by a preposition. Those clefts, however, were excluded from the analysis.

grammatical functions, we set up a comparison corpus of 200 randomly chosen non-clefted sentences from the same texts in which we found the clefts. Those sentences contained both main clauses and subordinate clauses, given that we found main clause clefts and subordinate clefts.

We analyzed the data in two ways: (i) We determined the relative frequencies of subjects and objects in the comparison corpus by counting all of their occurrences. Those frequencies were compared to the observed relative frequencies f_{cleft} of subject versus object clefts in the cleft sample. This method assumes that every grammatical argument is equally likely to be clefted, independent of the sentence it belongs to. So it ignores the fact that various grammatical arguments are unevenly distributed in sentences. (ii) For the second analysis, it is assumed that each sentence is equally likely to become a cleft. As sentences can have different numbers of arguments, this means that arguments of different sentences can now have different probabilities to be clefted. For example, compare two sentences of the form S-V-O and S-V-O-O. Both sentences are equally likely to become clefts, but have a different number of grammatical arguments. If the first sentence is selected, the probability that the subject is clefted is 0.5, as there are only two grammatical arguments which can be clefted. It is not possible to cleft the verb. If the second sentence is selected, this probability drops to 0.33, as there are now three possible grammatical arguments to be clefted. We calculated the probability of being clefted for each subject and object in each sentence from the comparison corpus and calculated their average over all sentences p_{cleft} , which was then compared to f_{cleft} . Each of the approaches can be seen as a useful simplification because the aspects they ignore are independent of each other.

We annotated several other properties of each cleft and its context. It is generally assumed in literature that clefts have an existence presupposition and an exhaustivity inference of some sort. The following inferences would be predicted for the cleft in (1).

- a. Existence presupposition: Somebody suggested an international re-orientation of the foundation on the 50th anniversary.
- b. Exhaustivity inference: Nobody other than the former president Roman Herzog suggested an international re-orientation of the foundation on the 50th anniversary.

The analysis of those inferences in our corpus, however, turned out to be unfeasible, as the inter-annotator agreement was too low for these features. The property ‘contrast’ was especially difficult to annotate since the notion is not well-defined. Following Repp (2010), we did annotate some categories related to contrast, such as the existence of explicitly mentioned alternatives and their

negation. Those categories did not seem to play a role in our sample. Taking an intuitive point of view, however, the cleft in example (4) clearly constitutes a contrast between Tony Blair and Gordon Brown. For the purpose of annotation, however, it is not obvious as to how to operationalize contrast in this and similar examples. Repp's (2010) criteria do not apply.

- (4) Tony Blair, zuletzt in seiner Partei geradezu verhasst, hat sensationelle drei Wahlsiege errungen; es war sein nach links rückender Nachfolger Gordon Brown, der abgewählt wurde.

Gordon Brown, who voted out was.

'Tony Blair, who was virtually hated in his party lately, achieved three sensational election victories; it was his left-moving successor Gordon Brown who was voted out of office.'

(Z10/OKT.03679 Die Zeit [Online-Ausgabe], 07.10.2010; Abschied vom Klassenfeind)

Since these properties did not seem to play a role in our sample, we will ignore them in our analysis.

3.2 Results

Table 1 presents the absolute numbers of subjects and objects found in the comparison sample and in the cleft pivots of the cleft sample.

Table 1: Absolute numbers n_{cleft} for the cleft sample and n_{comp} for the comparison corpus.

	n_{cleft}	n_{comp}
Subjects	249	192
Objects	24	93

Both approaches described above yield that subject clefts occur significantly more often than object clefts even with respect to the general frequency of subjects and objects. For approach (i), we tested the relative frequencies f_{cleft} of subjects and objects from the cleft sample and the relative frequencies f_{comp} from the comparison corpus for significant deviation using a χ^2 -test. The frequencies are displayed in Table 2. The test shows that subject clefts are significantly more frequent in the cleft sample ($p < 0.01$).

Table 2: Frequencies of subjects and objects in the cleft sample (f_{cleft}) and the comparison sample (f_{comp}), and the average probability (p_{cleft}) of subjects and objects in the comparison sample.

	f_{cleft}	f_{comp}	p_{cleft}
Subjects	0.91	0.67	0.76
Objects	0.09	0.33	0.24

For approach (ii), we tested f_{cleft} and the average probabilities p_{cleft} of subjects and objects from the comparison corpus (also displayed in Table 2) for significant deviation using a t-test. This test shows that subject clefts are significantly more frequent in the cleft sample than predicted by p_{cleft} ($p < 0.01$).

One natural explanation of the data could be that subjects are just clefted more frequently because of other properties that often co-occur with subjecthood, such as agentivity and animacy. After comparing these properties for subjects in the comparison corpus and subjects in the cleft pivots of our cleft sample, we can rule out this objection. Table 3 and 4 show that both samples demonstrate the same distribution for animate/non-animate and agentive/non-agentive subjects. A χ^2 -test yielded a p-value of $p = 0.39$ for animacy and $p = 0.56$ for agentivity. Hence, those properties do not seem to be the crucial ones.

Table 3: Absolute numbers (and %) of (in-) animate subjects in the cleft sample n_{cleft} and the comparison corpus n_{comp} .

	n_{cleft}	n_{comp}
Subjects [+animate]	117 (47%)	97 (52%)
Subjects [-animate]	132 (53%)	91 (48%)

Table 4: Absolute numbers (and %) of (non-) agentive subjects in the cleft sample n_{cleft} and the comparison corpus n_{comp} .

	n_{cleft}	n_{comp}
Subjects [+agent]	81 (33%)	71 (37%)
Subjects [-agent]	161 (67%)	123 (63%)

4 Discussion

Our results indicate a higher frequency of subject clefts as opposed to object clefts in German. We follow a line of argumentation similar to what is proposed by Féry (2001) and Szendrői (1999). We take focus to be a semantic notion (Krifka 2008). The focused element is syntactically marked by an F-feature (Rooth 1992) which is realized at the phonological form with an A-accent (Bolinger 1958). Contrary to French, in spoken German it is generally possible to mark focus by intonation in any position (including the subject position), as indicated in (5). However, this is different when it comes to written German. Here, the reader cannot identify the focus by referring to intonation but needs to rely on other cues provided in the text. The overt question in (5) could be such a cue.

- (5) Wer hat einen Apfel gegessen? – NINA hat einen Apfel gegessen.
 Who has an apple eaten? – NINA has an apple eaten.
 ‘Who ate an apple? – NINA ate an apple.’
- (6) Nina hat einen APFEL gegessen.
 Nina has an APPLE eaten.
 ‘Nina ate an APPLE.’
- (7) Es ist NINA, die einen Apfel gegessen hat.
 It is NINA who an apple eaten has.
 ‘It is NINA who ate an apple.’
- (8) Nina hat das Buch dem MANN geschenkt.
 Nina has the book the MAN given.
 ‘Nina gave the book to the MAN.’

If the context does not provide such a cue, the reader is likely to rely on her knowledge of where the default focus accent lies, that is, as in French, at the right edge of a phonological phrase.⁴ In many cases, the default intonation results in the object (not the subject) receiving highest prominence, as in (6). Hence, the object would be identified as the focus. Furthermore, objects can be scrambled into a position where they receive the default focus accent. In (8),

4 This does not imply that the reader constructs an actual prosodic-phonological representation for the written text although some studies would support that (for an overview of related research see Leininger 2014). For our argument to hold, it suffices that the reader just uses her knowledge of where the accent is ‘usually’ assigned.

for instance, the indirect object NP *dem Mann* ('the man') is scrambled to the end of the phonological phrase, where it is focused by default. In order to disambiguate focus-marking on the subject in written German, special marking is helpful (DeVeugh-Geiss et al. 2015). The cleft construction puts the subject into a position where it receives highest prominence by default (Szendrői 1999) and, thus, gives the reader a cue to identify the subject as the focus (see example (7)).

Following Féry (2001) and DeVeugh-Geiss et al. (2015), we argue that a cleft construction in its function of marking focus appears more often with subjects since there are other additional options to mark focus on objects, such as default focus accent or scrambling, which are inapplicable to subjects. In their base position, subjects do normally not receive a default accent. Furthermore, subjects are unlikely to be scrambled in order to be focused.

The question is now whether disambiguating subject focus is indeed the main motivation for using a cleft. Literature on clefts has mentioned several other features of clefts that might be worth considering, e.g., exhaustivity or the existential presupposition as explained in Section 3.1. Firstly, our annotation data did not provide clear evidence for the relevance of those features. Our argument is further strengthened by the observation that clefts are hardly ever used in spoken German.⁵ An account just based on the existential presupposition and/or exhaustivity cannot explain the difference between the frequency of clefts in spoken and written German. Neither the existential presupposition nor the exhaustivity inference seem plausible to have an effect on the frequency of clefts in general. In particular, there is no reason why those properties should be more developed in written than in spoken German.

Our analysis of clefts as devices to shift prominence away from the default, in contrast, predicts there to be fewer clefts in spoken German. In spoken German there is simply no need for a cleft construction since focus can always be disambiguated using intonation by marking an element in-situ, as in (5). This option is missing in written German, which leads to more clefts in written German. Our analysis is nevertheless compatible with assigning an exhaustivity inference and an existential presupposition to clefts, but those features are not assumed to constitute the main motivation for using a cleft.

5 Even though we did not conduct a quantitative study about the frequency of clefts in spoken German, our informants and the native speaker judgments of the authors support the low frequency of clefts in spoken German.

5 Conclusion

From our data set, we can conclude that subjecthood is the main factor determining the use of clefts, possibly due to the wish of the author to give cues for unambiguously identifying the focused element in the sentence. This is in line with the observation that subject clefts occur more often than object clefts since German has other ways of disambiguating focus for objects, e.g., default intonation and scrambling. Our approach is also capable of predicting a difference between spoken and written German.

Some issues are left open here and will need further research. So far, our reasoning only works for focus-background clefts, but should be extended to also cover topic-comment clefts. Moreover, the role of contrast should be operationalized for annotation or further analyzed using other methods.

References

- Bolinger, Dwight L. 1958. A Theory of Pitch Accent in English. *Word*, 14(2-3), 109-149.
- Carter-Thomas, Shirley 2009. The French c'est-cleft: Function and frequency. In D. Banks (ed.), *La linguistique systématique fonctionnelle et la langue française*, 127-157. Paris : L'Harmattan.
- Destruel, Emilie. 2012. The French c'est-cleft: An empirical study on its meaning and use. *Empirical issues in Syntax and Semantics*, 9, 95-112.
- DeVeugh-Geiss, Joseph P., Malte Zimmermann, Edgar Onea and Anna-Christina Boell. 2015. Contradicting (not-)at-issueness in exclusives and clefts: An empirical study. *Semantics and Linguistic Theory*, 25, 373-393.
- Féry, Caroline. 2001. Focus and phrasing in French. In C. Féry and W. Sternefeld (eds.), *Audiatur Vox Sapientia. A Festschrift for Arnim von Stechow*, 153-181. Berlin: Akademie Verlag.
- Hartmann, Katharina & Zimmermann, Malte. 2007. In place-out of place? Focus in Hausa. In K. Schwabe and S. Winkler (eds.), *On Information Structure, Meaning and Form: Generalizations across languages*, 365-403. Amsterdam: John Benjamins Publishing Company.
- Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica*, 55(3-4): 243-276.
- Lambrecht, Knut 2001. A framework for the analysis of cleft constructions. *Linguistics*, 39(3): 463-516.
- Leinenger, Mallorie. 2014. Phonological coding during reading. *Psychological Bulletin*, 140(6): 1534-1555.
- Reinhart, Tanya 1995. Interface strategies. *OTS working papers in linguistics*.

- Repp, Sophie 2010. Defining 'contrast' as an information-structural notion in grammar. *Lingua*, 120(6): 1333–1345.
- Roland, Douglas, Frederic Dick and Jeffrey L. Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57(3): 348–379.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural language semantics*, 1(1): 75–116.
- Skopeteas, Stavros and Gisbert Fanselow. 2010. Focus types and argument asymmetries: A crosslinguistic study in language production. In C. Breul and E. Göbbel (eds.), *Contrastive information structure*, 169–197. Amsterdam: John Benjamins Publishing Company.
- Szendrői, Kriszta. 1999. A stress-driven approach to the syntax of focus. *UCL Working Papers in Linguistics*, 11, 545–573.