Methodological Considerations on Testing Argument Asymmetry in German Cleft Sentences

Swantje Tönnis, Lea M. Fricke, Alexander Schreiber (Georg-August-Universität Göttingen)

Introduction: We present a corpus study with the aim of contributing to a better understanding of the factors that facilitate the use of *it*-clefts in German. We analyzed crucial properties of clefts and their contexts. In this poster presentation, 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, one can distinguish between subject clefts as in (1), and non-subject clefts.

(1) Es ist Peter, der Maria liebt.
It is Peter, who$_{\text{NOM.SG}}$ Maria loves (*'It is Peter who loves Maria.'*)

For several languages, it has been claimed that subject clefts are more frequent than non-subject clefts (see Carter-Thomas 2009, Roland et al. 2007, and Skopeteas & Fanselow 2010). We tested this claim for German clefts. However, we used a more complex method than earlier studies on other languages. We did not only provide the frequencies of (non-)subject clefts, but also compared those frequencies to the general frequency of (non-)subjects. 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.

Methods: We drew sample of 300 random clefts from the DeReKo corpus of written German and annotated the grammatical function of each cleft relative pronoun. Moreover, we set up a comparison corpus of 200 non-clefted sentences from the same texts in which we found the clefts in order to capture the frequency of certain grammatical functions in general. We analyzed the data in two ways: (i) We determined the relative frequencies of (non-)subjects in the comparison corpus by counting all of their occurrences in order to compare them to the observed relative frequencies f_{cleft} in the cleft sample, ignoring that various grammatical arguments are unevenly distributed in CPs. (ii) We calculated the probability to be clefted for each (non-)subject in each sentence from the comparison corpus and calculated the average p_{cleft} of the probabilities of a (non-)subject to be clefted over all sentences and compared them to f_{cleft}. This approach rests on the idealized assumption that each CP is equally likely to become a cleft. Each approach can be seen as a useful simplification because the aspects they ignore are independent of each other.

Results: Both approaches yield that subject clefts occur significantly more often than non-subject clefts even with respect to the general frequency of subjects/non-subjects. Table 1 displays the absolute numbers and Table 2 and Table 3 display the relative frequencies for each approach. For (i), we used a χ^2-test, for (ii) a t-test (both p-values of $p<0.01$).

Discussion: Our results indicate a higher frequency of subject clefts in German. An explanation is 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 non-subjects, such as prosodic prominence or movement, which are inapplicable to subjects (Féry 2001, DeVeaugh-Geiss et al. 2015). The cleft construction puts the subject into a position where it receives prominence (Szendröi 1999) and, thus, gives the reader a clue of the intended intonation. We also tested the importance of contrast, as a motivation to use a cleft, by annotating several other categories for our clefts (including contexts) that could possibly relate to contrast, such as negation or the availability of explicit alternatives to the pivot (Repp 2010). However, those categories did not turn out to play a role in our sample.

Conclusion: From our data set, we can so far only conclude that subjecthood is the main factor determining the use of clefts, possibly due to the wish of the speaker to give cues for the prosodic prominence of the argument functioning as subject. This is in line with the observation that subject clefts occur more often than non-subject clefts since German has other ways of making a non-subject NP prominent, e.g. default intonation and movement to the prefield.

Tables

	n_{cleft}	n_{comp}
Subjects	249	191
Non-Subjects	51	274

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

	f_{cleft}	f_{comp}
Subjects	0.83	0.41
Non-Subjects	0.17	0.59

Table 2: Results for approach (i).

	f_{cleft}	p_{cleft}
Subjects	0.83	0.51
Non-Subjects	0.17	0.49

Table 3: Results for approach (ii).

Source

Das Deutsche Referenzkorpus DeReKo, http://www.ids-mannheim.de/kl/projekte/korpora/, Institut für Deutsche Sprache, Mannheim

References

CARTER-THOMAS, S. (2009). The french c'est-cleft: Function and frequency. In Banks, D., editor, *La linguistique systémique fonctionnelle et la langue française*, pages 127–157. L'Harmattan, Paris.

DEVEAUGH-GEISS, J., ZIMMERMANN, M., ONEA, E., AND BOELL, A.-C. (2015). Contradicting (not-)at-issueness in exclusives and clefts: An empirical study. *Semantics and Linguistic Theory*, 25:373–393.

FÉRY, C. (2001). Focus and phrasing in french. In Fery, C. and Sternefeld, W., editors, *Audiatur Vox Sapientia. A Festschrift for Arnim von Stechow*, pages 153–181. Akademie Verlag, Berlin.

REPP, S. (2010). Defining 'contrast' as an information-structural notion in grammar. *Lingua*, 120(6):1333–1345.

ROLAND, D., DICK, F., AND ELMAN, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57(3):348–379.

SKOPETEAS, S. AND FANSELOW, G. (2010). Focus types and argument asymmetries: a crosslinguistic study in language production. In Breul, C. and Göbbel, E., editors, *Contrastive information structure*, pages 169–197, Benjamins, Amsterdam, Philadelphia.

SZENDRÖI, K. (1999). A stress-driven approach to the syntax of focus. *UCL Working Papers in Linguistics*, 11:545–573.