

# Focus in German: Towards a Corpus-based study

Kordula De Kuthy and Detmar Meurers

Seminar für Sprachwissenschaft, Universität Tübingen

{kdk, dm}@sfs.uni-tuebingen.de

## 1 Motivation and Background

The nature of the integration of a sentence into the discourse can provide an explanation for constraints previously stipulated in syntax (cf., e.g., Cook, 2001; De Kuthy, 2002; De Kuthy and Meurers, 2003). However, to explore this line of research, one needs an explicit understanding and representation of the interaction of syntax and information structure.

English and German are so-called intonation languages where information structuring is signaled by the intonation of an utterance, including pitch accents. The absence or presence of an accent thus is an indicator of the discourse function of a particular constituent in a sentence. The research investigating the interaction of syntax, information structure, and intonation has traditionally been theoretically driven, with the syntactic F-marking approach of Selkirk (1995) serving as a frequently used foundation. At the same time, recent work mostly driven by pragmatic and semantic considerations (cf., e.g., Büring, 2006; Roberts, 2006; Kadmon, 2006) has questioned the very foundation of such an approach. This includes the claim that focus projection as the fundamental means of connecting the focus exponent (pitch accent) and the semantically interpreted focus element is not needed. Interestingly, the different approaches do not just differ in terms of their theoretical interpretation, but they also make claims about a fundamentally different empirical landscape. Roberts (2006) and Kadmon (2006) assume significantly more pitch accents than have previously been assumed, and claim that focus projection is not needed. Büring (2006), on the other hand, claims that focus projection is in principle always possible, from any element in the sentence, thereby negating the need for the theory of focus projection constraints traditionally assumed.

The field thus is in a situation where drastically different theoretical interpretations are based on dramatically different, contradictory empirical assumptions. Unfortunately, there is only little published research on the empirical evidence relating to focus projection. The published work we are aware of involves comprehension studies (Gussenhoven, 1983; Birch and Clifton, 1995; Welby, 2003; Féry, 1993). These studies generally support the existence of focus projection.

In this abstract reporting work in progress, we contribute to the empirical evidence relating to focus projection in German by studying production. We explore where spoken language corpora can provide empirical evidence for or against the different conceptualizations of focus projection. Complementing the work on corpora collected by Project D2 of the SFB 632 using the QUIS questionnaire eliciting data from informants, we investigate authentic utterances found in already existing, intonationally annotated corpora.

## 2 Focus Projection

Focus projection rules are traditionally viewed as the link between the word carrying the (nuclear) pitch accent and the part of the utterance interpreted as being in focus (new information in the discourse). Example (2) illustrates this, where the exact same sentence, with a pitch accent on “BAGgern” (accent on capitalized syllable), is assumed to be able to occur in the different contexts below, where the extent of the projected focus is marked by [[...]]F

- (1) *Marius hat ein Buch mit BAGgern bekommen.*  
 Marius has a book with diggers received  
 ‘Marius received a book on diggers (as a present).’
- (2) a. *Was für ein Buch hat Marius bekommen?* (What kind of a book did Marius receive?)  
*Marius hat ein Buch [[mit BAGgern]]F bekommen.*  
 b. *Was hat Marius bekommen?* (What did Marius receive?)  
*Marius hat [[ein Buch mit BAGgern]]F bekommen.*  
 c. *Wie war Weihnachten für Marius?* (So how was Christmas for Marius?)  
*Marius [[hat ein Buch mit BAGgern bekommen]]F.*  
 d. *Was war los?* (What happened?)  
*[[Marius hat ein Buch mit BAGgern bekommen]]F.*

As mentioned in the introduction, some authors have questioned the very existence of such focus projection (e.g., Roberts, 2006; Kadmon, 2006). So we next turn to the question what evidence can be found in a corpus about this issue.

## 3 The IMS Radionews Corpus

Before diving into the specifics of the corpus used, let us be clear that corpus data needs to be interpreted with care. The fact that a particular type of example was found in a corpus, does not necessarily mean that it is a systematic instance which needs to be licensed by linguistic theories. Similarly, the absence of a particular type of example in a corpus does not mean that it should not be licensed, given that following Zipf’s law (Zipf, 1936) most things will occur only rarely and corpora are

limited in size. Nevertheless, corpus data can provide important empirical insights for theoretical linguistic analysis (cf., e.g., Meurers, 2005).

We base our study on the IMS Radionews Corpus (Rapp, 1998), one of the few intonationally annotated corpora of German. It includes recordings of radio broadcasts on the Deutschlandfunk for a total length of 1 hour and 26 minutes, amounting to 514 sentences. The corpus preparation included manual segmentation into news stories, orthographic transliteration, automatic word alignment, phonetic transcription, and manual prosodic labeling with ToBI (Beckman and Pierrehumbert, 1986).

Searching for the relevant focus projection patterns in a corpus is made significantly easier if one can refer to constituents, yet the IMS Radionews Corpus is not syntactically annotated. We therefore parsed the corpus with the Berkley parser (Petrov and Klein, 2007). While the resulting annotation is not perfect, we found that it is of high enough quality to search for the relevant patterns with sufficient precision and recall. Following syntactic annotation, we converted the corpus into TiGer-XML format, so that it can be browsed and searched using the TiGerSearch tool (Lezius, 2002). The converted corpus includes the orthographic transcription, the phonetic transcription, the ToBI annotation, and the syntactic analysis.

## **4 Some Findings**

In this abstract we can only provide a brief overview of the results which are more fully discussed in the talk. The first observation is that one does find examples which seem to be instances of focus projection. For example, in the example shown on the page following the references, the strongest accent falls on the last element of the PP (following the Nuclear Stress Rule of Chomsky and Halle, 1968 for English; Jacobs, 1988, p. 124 for German) but the entire constituent is focused. On the other hand, we found many examples with significantly more accents than are traditionally assumed by syntactic theories of focus projection, with some examples carrying pitch accents on almost all of the words. We also found examples with accents that occur in positions that are unexpected for standard theories of focus projection. And finally, there seems to be significant variation in the prosodic realization, which we are able to study given that the same news items appear in several news announcements in the corpus.

## **References**

- Beckman, M. and J. Pierrehumbert (1986). Intonational Structure in Japanese and English. *Phonology Yearbook* 3.
- Birch, S. and C. Clifton, Jr (1995). Focus, Accent, and Argument Structure: Effects on Language Comprehension. *Language and Speech* 38(4), 365–391.

- Büring, D. (2006). Focus Projection and Default Prominence. In V. Molnár and S. Winkler (eds.), *The Architecture of Focus*, Berlin: M. De Gruyter, pp. 321–346.
- Chomsky, N. and M. Halle (1968). *The Sound Pattern of English*.
- Cook, P. (2001). *Coherence in German: An Information-Structural Account*. Ph.D. thesis, University of Manchester.
- De Kuthy, K. (2002). *Discontinuous NPs in German. A Case Study of the Interaction of Syntax, Semantics and Pragmatics*. Stanford, CA: CSLI Publications.
- De Kuthy, K. and W. D. Meurers (2003). The Secret Life of Focus Exponents, and What It Tells Us about Fronted Verbal Projections. In S. Müller (ed.), *Proceedings of the Tenth Int. Conference on HPSG*. Stanford, CA: CSLI Publications.
- Féry, C. (1993). *German Intonational Patterns*. Tübingen: Max Niemeyer Verlag.
- Gussenhoven, C. (1983). Testing the reality of focus domains. *Language and Speech* 26, 61–80.
- Jacobs, J. (1988). Fokus-Hintergrund-Gliederung und Grammatik. In H. Altmann (ed.), *Intonationsforschungen*, Tübingen: Max Niemeyer Verlag, pp. 89–134.
- Kadmon, N. (2006). Some Theories of the Interpretation of Accent Placement. Handout for Talk at The Ohio State University, October 19, 2006.
- Lezius, W. (2002). Ein Suchwerkzeug für syntaktisch annotierte Textkorpora. Ph.D. thesis, IMS, University of Stuttgart. AIMS report, vol 8, nr 4.
- Meurers, W. D. (2005). On the use of electronic corpora for theoretical linguistics. Case studies from the syntax of German. *Lingua* 115(11), 1619–1639.
- Petrov, S. and D. Klein (2007). Improved Inference for Unlexicalized Parsing. In *HLT/ACL 2007*. pp. 404–411.
- Rapp, S. (1998). *Automatisierte Erstellung von Korpora für die Prosodieforschung*. Ph.D. thesis, Univ. Stuttgart.
- Roberts, C. (2006). *Resolving Focus. Sinn und Bedeutung* 11. Barcelona, Spain.
- Selkirk, E. (1995). Sentence Prosody: Intonation, Stress, and Phrasing. In J. A. Goldsmith (ed.), *The Handbook of Phonological Theory*, Oxford: Basil Blackwell.
- Welby, P. (2003). Effects of Pitch Accent Position, Type, and Status on Focus Projection. *Language and Speech* 46(1).
- Zipf, G. K. (1936). *The Psycho-Biology of Language*. London: Routledge.

