Focus in German: Towards a Corpus-Based Study

Kordula De Kuthy and Detmar Meurers

seminar für sprachwissenschaft

Eberhard Karls Universität Tübingen

Motivation

- Nature of integration of a sentence into discourse can provide explanations for constraints stipulated in syntax (De Kuthy, 2002; De Kuthy & Meurers, 2003).
  - To explore this line of research, we need an explicit representation of information structure and its interaction.
- Prosody plays important role in constraining possible integration of a sentence into discourse.
- Research relating syntax, information structure, and intonation predominantly theoretically driven with partly contradictory empirical assumptions.

⇒ Investigate empirically by looking at the intonation of naturally occurring sentences in a larger discourse.

- Pilot study using intonationally annotated corpora

IMS Radionews Corpus

- IMS Radionews Corpus (Rapp, 1998): Recordings of radio broadcasts on Deutschlandfunk:
  - 18 consecutive news broadcasts (28.07./21.11.95) Total length: 1h 26min, 514 sent. (≈ 10 sec/sent)
- Preparation of the corpus involved:
  - manual segmentation into news stories
  - orthographic transliteration
  - automatically word alignment
  - manual prosodic labeling with ToBI

Tones and Break Indices (ToBI)

- ToBI: system for transcribing the intonation patterns and other aspects of the prosody of English.
  - Based on autosegmental-metrical approach to intonation Beckman & Pierrehumbert (1986), for German: Mayer (1995); Grice et al. (2002)
- Perceived intonation contour described in terms of high (H) and low (L) targets in the local pitch range.
  - accented syllables, for English:
    - H*,L*, or bitonal: H*+L, H+L*, L*+H, L+H*, H*+H
    - The * marks the tone on the accented syllable.
  - intonational boundaries (break values: 0–4):
    - intermediate boundary (0–3): H^-, L^-
    - full boundary (4): L% or H%

Relating Syntax, Intonation & Discourse

- Focus projection rules determine the focus projection potential of a pitch accent in a syntactic tree.
  - Popular syntax-driven approach to focus projection: Selkirk (1995) and many variants
- Strongest accent is generally constrained to fall on the last element (e.g., Nuclear Stress Rule of Chomsky & Halle, 1968; for German: Jacobs, 1988)

(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*$]\!]_F$.

The Open Issue

- Several authors have recently questioned whether there is an interesting constraining relation between syntax, intonation, and information structure:
  - Büring (2006) any accent within a phrase can project focus, i.e., focus can always project
  - Kadmon (2006) & Roberts (2006): focus never projects (new elements → accented; unaccented & focused → given/retrievable/expectable)
- Very little published empirical research on focus projection, all on perception (Gussenhoven, 1983; Birch & Clifton, 1995; Welby, 2003; Féry, 1993).
  - substantial evidence for the existence of some focus projection, i.e., for certain pairs of verbs with object-NP arguments in English
- More evidence, especially on production is needed.
  - In which constructions
  - can what kind of elements
  - be accented with which type of accents
  - and project focus how far?

Pilot Study using IMS Radionews Corpus

- IMS Radionewscorpus not syntactically annotated.
  - We parsed the corpus with the Berkley parser (Petrov & Klein, 2007) to obtain syntactic analysis.
    - Thanks to Adriane Boyd
- We converted the corpus into TiGer-XML including
  - orthographic transcription
  - phonetic transcription
  - ToBI annotation
  - syntactic analysis

  so that it can be browsed and searched using TiGerSearch tool (Lezius, 2002).

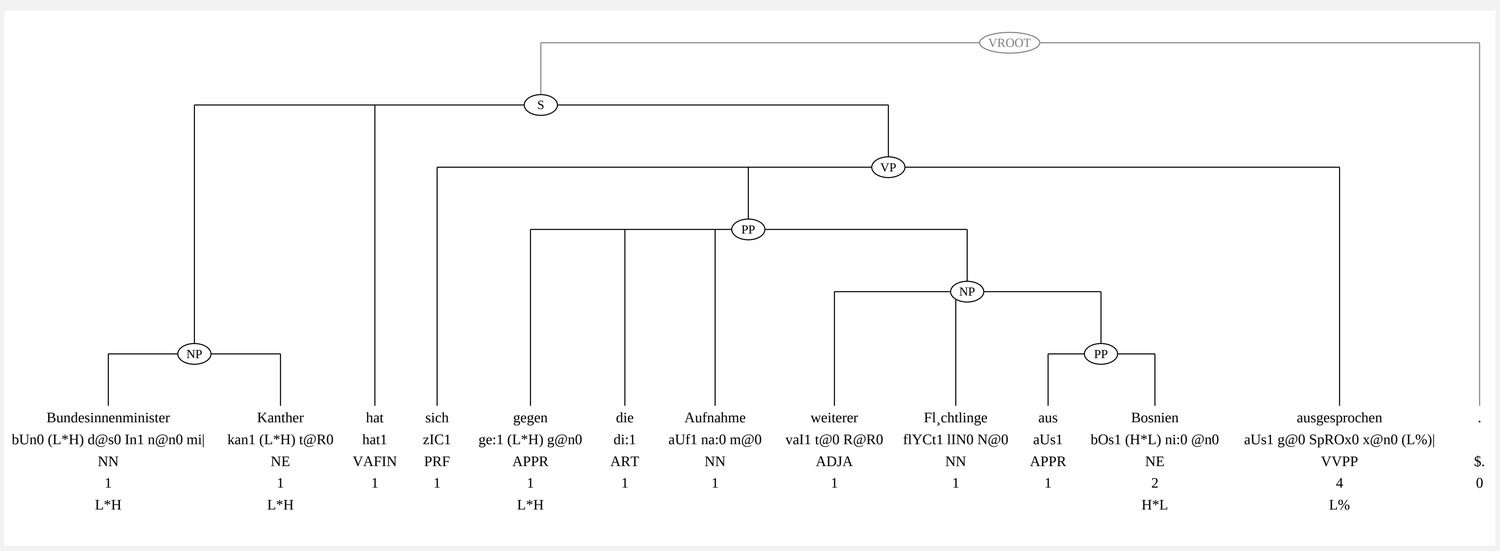
Observations

- Some examples show **focus projection with the traditionally assumed accent pattern**.
- In others, one finds **more accents** than assumed by traditional theories of focus projection.
- A number of **accents in unexpected positions** occur, given standard theories of focus projection.
- There is **significant variation**, even when the same information in the same context is reported.

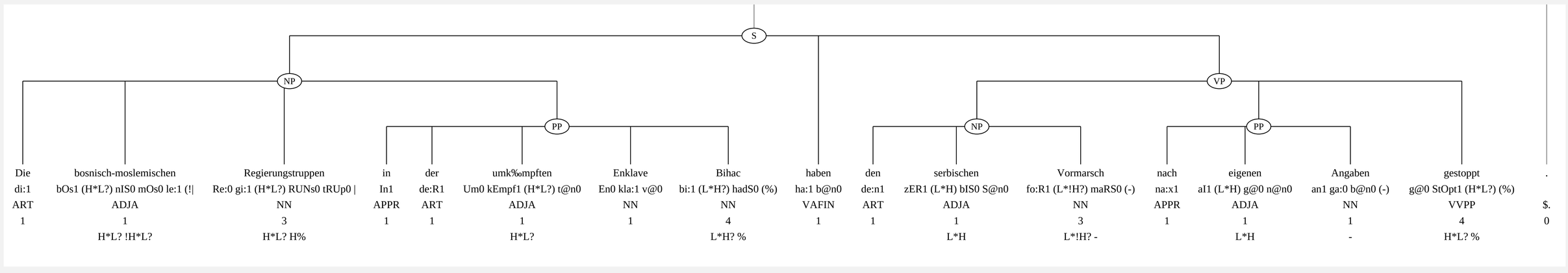
Outlook

- Investigate status of additional, prenuclear accents often occurring in NPs from which focus projects.
- Investigate more spontaneous spoken language instead of read news, with its special discourse structure and potentially unusual intonation.
  - For news, difficult to determine what information is accessible in extra-linguistic context at the time.
  - Potential candidate: The Verbmobil (VM) corpus of appointment scheduling dialogues. However:
    - Many short, incomplete utterances and nature of discourses very limited.
    - Hardly any data, all first phase, annotated with pitch accents (later: only intonational phrasing).

IMS Radionews Corpus Example for Focus Projection in an NP



IMS Radionews Corpus Example for Multiple Accents in an NP



References

Beckman, M. & J. Pierrehumbert (1986). Intonational Structure in Japanese and English. *Phonology Yearbook* 3, 255–309.

Birch, S. & C. Clifton, Jr (1995). Focus, Accent, and Argument Structure: Effects on Language Comprehension. *Language and Speech* 38(4).

Büring, D. (2006). Focus projection and default prominence. In V. Molnár & S. Winkler (eds.), *The Architecture of Focus*, Berlin: Mouton De Gruyter.

Chomsky, N. & M. Halle (1968). *The Sound Pattern of English*. Harper & Row.

De Kuthy, K. (2002). *Discontinuous NPs in German — A Case Study of the Interaction of Syntax, Semantics and Pragmatics*. CSLI Publications.

De Kuthy, K. & 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*. CSLI Publications.

Féry, C. (1993). *German Intonational Patterns*. Tübingen: Niemeyer.

Grice, M., S. Baumann & R. Benzmüller (2002). German Intonation within the Framework of Autosegmental-Metrical Phonology. In S.-A. Jun (ed.), *Prosodic Typology and Transcription: A Unified Approach*, Oxford: Oxford University Press.

Gussenhoven, C. (1983). Testing the reality of focus domains. *Language and Speech* 26.

Jacobs, J. (1988). Fokus-Hintergrund-Gliederung und Grammatik. In H. Altmann (ed.), *Intonationsforschungen*, Tübingen: Niemeyer.

Kadmon, N. (2006). Some Theories of the Interpretation of Accent Placement. Handout, colloquium at OSU, October 19, 2006.

Lezius, W. (2002). Ein Suchwerkzeug für syntaktisch annotierte Textkorpora. Ph.D. thesis, IMS, University of Stuttgart.

Mayer, J. (1995). *Transcribing German Intonation — The Stuttgart System*. Tech. rep., Universität Stuttgart. http://www.ims.uni-stuttgart.de/phonetik/joerg/labman/STGTsystem.html.

Petrov, S. & 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, University of Stuttgart.

Roberts, C. (2006). Resolving Focus. Conference abstract. Sinn und Bedeutung 11. Barcelona, Spain.

Selkirk, E. (1995). Sentence Prosody: Intonation, Stress, and Phrasing. In *The Handbook of Phonological Theory*, Oxford: Basil Blackwell, chap. 16.

Welby, P. (2003). Effects of Pitch Accent Position, Type, and Status on Focus Projection. *Language and Speech* 46(1), 53–81.